Music Perception
Summer 2005, Vol. 22, No. 4, 583–611

Relationship Between Complexity and Liking as a Function of Expertise

MARK G. ORR
Albert Einstein College of Medicine

STELLAN OHLSSON
University of Illinois at Chicago

The inverted-U hypothesis has much empirical support in the field of experimental aesthetics. This hypothesis predicts that moderately complex art objects should be preferred over very simple or very complex art objects. Although it is tacitly believed that this hypothesis applies to experts, the literature does not contain any convincing studies that demonstrate this as fact. The present study addresses this issue. Professional jazz and bluegrass musicians rated the complexity of and their liking for short, but complete, jazz and bluegrass improvisations. Complexity and liking were operationalized by subjects' judgments on seven-point Likert-like scales. Regressing liking onto complexity did not reveal any evidence for an inverted-U relation for experts. Moreover, no relationship between liking and complexity was found for the jazz musicians; a negative relation was found for the bluegrass musicians, but only when listening to the bluegrass improvisations. Furthermore, by comparing the expert data with a reanalysis of nonexpert data collected in a previous, but identical study, we propose as a first approximation that musical expertise dissolves the relationship between liking and complexity.

Received August 24, 2001, accepted March 27, 2004.

AESTHETIC judgment occurs in all human cultures and is likely to be a fundamental function of the human mind. The determinants of aesthetic response are still not fully understood. In this article, we approach the question of the determinants of aesthetic judgments of music by studying the interaction between complexity and liking in expert musicians.

Address correspondence to Mark G. Orr, Department of Neuroscience, Kennedy 825, Albert Einstein College of Medicine, 1410 Pelham Pkwy. S., Bronx, NY 10461. (e-mail: morr@aecom.yu.edu)

ISSN: 0730-7829, electronic ISSN: 1533-8312. Please direct all requests for permission to photocopy or reproduce article content to University of California Press's Rights and Permissions website, at www.ucpress.edu/journals/rights.htm.

The rationale for this approach is rooted both in experimental aesthetics and in research on expertise.

Equating aesthetic judgment with human evaluative ratings of liking or pleasantness has been common practice in experimental aesthetics. For example, pleasantness (Guilford, 1934; Washburn, 1911), agreeableness (Pierce, 1894), and liking (Martin, 1906, North & Hargreaves, 1995; Orr & Ohlsson, 2001) are common dependent variables used when investigating aesthetic behavior. Furthermore, aesthetic judgment, in this line of research, is distinguished from highly emotional reactions to art (e.g., as studied by Gabrielsson, 2001). Berlyne (1974) separated evaluative and internal-state scales because he assumed that each measured a different aspect of the human response to art. To simplify somewhat, the former measure liking of art; the latter measure the mood induced by art. Berlyne's distinction is similar to more recent work in music and emotion. Abeles and Chung (1996) distinguish between the following three responses to music: "affective," "mood/emotional," and "preference." McMullen (1996) has three types of what he calls affective/aesthetic behavior: "preference/hedonic tone," "mood/feeling tone," and "aesthetic response." Scherer and Zentner (2001) posit several facets of emotional meaning in music, of which some are relevant for this discussion: "preferences," "emotions," and "mood." Despite the labeling differences, liking/pleasantness judgments are agreed-upon behaviors that are (a) associated with music and (b) distinct from emotional responses.

The Inverted-U Hypothesis

In the 19th century, Fechner reformulated the principle of unity in diversity—from philosophical aesthetics—using the psychological notion of *avoidance of extremes* in his 1876 book, *Vorschule der Aesthetik* (Arnheim, 1986). Aesthetic pleasure results from avoiding the extreme values of some stimulus variable. Although Fechner's own observations were inconclusive, his experimental approach stimulated empirical research on aesthetic judgment (Cohn, 1894, as cited in Chandler, 1934; Minor, 1909 as cited in Chandler, 1934; Pierce, 1894, 1896; Washburn, 1911). In fact, his idea of avoidance of extremes has dominated the field of experimental aesthetics up to the present (Arnheim, 1986; but see Martindale, 1984, 1988 for a different theoretical basis).

Although, in principle, the notion of avoidance of extremes applies to any stimulus variable, considerable attention has been paid to *stimulus complexity*. The extremes to be avoided are simplicity and chaos, with beauty striking a balance between the two. Although Fechner (as cited by Eysenck, 1968) attempted to define stimulus complexity (e.g., by varying the sides and angles of polygons), Birkhoff's (1933) work put stimulus complexity at the forefront of empirical aesthetics and served as the impetus for early research (e.g., Davis, 1936; Eysenck, 1941, 1942).

The principle of avoidance of extremes predicts that aesthetic response should vary with stimulus complexity as an inverted-U function. Both low and high levels of complexity generate low aesthetic responses; maximum liking occurs at intermediate complexity levels. This prediction is commonly referred to as the *inverted-U hypothesis* (e.g., Berlyne, 1971). The term refers both to an empirical regularity—that liking, when plotted as a function of complexity, exhibits an inverted-U function—and to the psychological mechanisms and processes that are hypothesized to generate aesthetic responses.

The empirical literature supports the existence of an inverted-U relation between liking and complexity for both real and artificial stimuli (Berlyne, 1974; Crozier, 1974; Dorfman, 1965; Dorfman & McKenna, 1966; Eckblad, 1963, 1964; Heyduk, 1975; Kamman, 1966; McMullen & Arnold, 1976; Munsinger & Kessen, 1964; North & Hargreaves, 1995, 1996a, 1996b; Radocy, 1982; Rump, 1968; Vitz, 1966a, 1966b; Walker, 1980; Wohlwill, 1968). However, the support is not unanimous. The relationship between complexity and liking has been found to be positive-linear (e.g., Day, 1967; Walker, 1980) or negative linear (Smith & Melara, 1990). Our own work found evidence for an inverted-U relation only for bluegrass improvisations and not for jazz improvisations (Orr & Ohlsson, 2001). Despite these inconsistencies, the inverted-U hypothesis is widely accepted (North & Hargreaves, 1995, 1996a).

The Inverted-U Hypothesis and Expertise

The prevailing conception in experimental aesthetics is that increased experience reduces perceived complexity and thereby shifts the apex of the inverted-U function toward greater complexity (e.g., Berlyne, 1971; Davies, 1978; Dember & Earl, 1957; Munsinger & Kessen, 1964; North & Hargreaves, 1995; Walker, 1980). A tentative prediction, then, is that the shape of the relation should be invariant across levels of expertise. However, the prediction is somewhat more complicated, as is explained next.

The prediction of the effects of increased experience on the complexity/liking relationship depend on whether the independent variable is either objective or perceived complexity. An example of the former is the use of information-theoretic variables such as uncertainty and redundancy (e.g., Vitz, 1964). An example of the latter is using verbal ratings of perceived complexity (e.g., North & Hargreaves, 1995; Orr &

Ohlsson, 2001). If an objective complexity measure is used as the independent variable, then the complexity of the stimuli themselves does not change as a function of expertise in terms of how complexity is measured. Consequently, when liking is plotted as a function of an objective complexity measure, increased experience is predicted to displace the inverted-U curve toward the right, in the direction of higher complexity levels (Figure 1a).

In contrast, if a perceived complexity measure is used as the independent variable, then it is predicted that the complexity of the stimuli will become less as a function of expertise (again, in terms of how complexity is measured). Therefore, when liking is plotted as a function of perceived complexity, each individual point in the plot of the complexity/liking relation is predicted to be displaced to the left, toward lower complexity, as expertise increases. Furthermore, the inverted-U hypothesis predicts that optimal liking remains at a moderate level of perceived complexity for all levels of expertise. Thus, the effect of training should push individual stimuli leftward along the inverted-U curve and not push the whole curve leftward, toward a lower level of complexity (Figure 1b).

Only a handful of empirical studies have addressed the effect of experience and training on the shape of the inverted-U relation. Of the five studies we reviewed that employed an objective complexity measure, three supported the predicted right displacement of the inverted-U curve and two did not. Crozier (1974), Munsinger and Kessen (1964), and Smith and Melara (1990) found an increase in the preferred complexity for participants with greater levels of training. In contrast, Simon and Wohlwill (1968) and Vitz (1966a) did not find comparable differences between more and less trained individuals. The two studies we reviewed that used

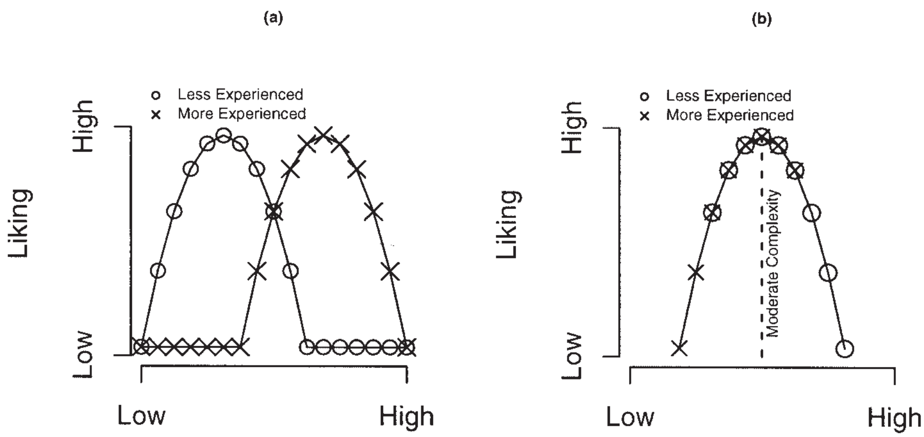

Fig. 1. A comparison of the hypothetical relationships between liking and (a) objective complexity versus (b) perceived complexity.

a perceived complexity measure did not support the prediction of leftward displacement of individual stimuli along the inverted-U curve. Neither North and Hargreaves (1995) nor Sinclair (as reported in Walker, 1980) found that training displaced the individual stimulus units toward less complexity. In fact, North and Hargreaves found that music training displaced the inverted-U higher on the perceived complexity scale; this is the prediction for objective complexity, not for perceived complexity. Moreover, Sinclair found that law students preferred a moderate or higher complexity level for art drawings and that art students preferred a moderate complexity level.

Taken together, these studies do not give a clear picture of how expertise affects the liking/complexity relation. The problem lies in the fact that either high-level experts were not used as participants and/or real art objects/music were not used as stimuli. This article directly addresses these two problems, as is explained next in more detail.

Question and Approach

The purpose of the study reported in this article was to clarify the relation between complexity and liking for experts in music. The question under investigation was whether the shape of the relation between complexity and liking approximates an inverted-U relation for experts.

Four methodological issues must be solved to investigate this question: (a) the ecological validity of the stimuli, (b) the procedure for manipulating complexity, (c) the choice of a predictor variable for liking, and (d) the level of expertise of the subjects. In the past, these methodological issues interacted to obscure the picture of the relationship between liking and complexity for experts listening to music. For example, some studies used true experts but not ecologically valid stimuli (e.g., Smith & Melara, 1990); some used ecologically valid stimuli, but not true experts (e.g., North & Hargreaves, 1995). Our responses to these four issues are as follows.

First, we used stimuli with high ecological validity for two reasons. The first reason is that it is easy to confound complexity and musicality. For example, Smith and Melara (1990) used variants of a simple nine-chord progression. They found a positive linear relationship between liking and what was called syntactic atypicality for the expert participants. However, their simplest stimuli were too simple to be perceived as music (e.g., I-V-I-IV-V-vi-IV-V-I, no inversions). Therefore, the experts might have preferred the more complex pieces because they approximated music whereas the simplest pieces did not. The problem of ecological validity was partly overcome by North and Hargreaves (1995), who used 30-second excerpts from popular music at various complexity levels. However, such excerpts do not have natural-sounding beginnings or endings. Furthermore, by using excerpts of popular music, North and Hargreaves may have confounded mere exposure effects (Moreland & Zajonc, 1976; Zajonc, 1968) with complexity. The second reason to use ecologically valid stimuli is that the expertise literature claims that the effects of training are highly specific to the particular domain (musical style) in which training occurs (e.g., Chi, Glaser & Farr, 1988; Ericsson & Smith, 1991). This excluded the use of artificial stimuli such as computer-generated tone sequences (Vitz, 1966a). To be able to assess the effects of training, we had to use musical stimuli that would engage the musical knowledge of our subjects.

In response to these observations, we employed two professional jazz and two professional bluegrass musicians to produce short improvisations that we used as stimuli in the study. By asking musicians to generate the stimuli, we ensured that the latter would be perceived as belonging to the relevant musical domain (bluegrass or jazz) by other musicians in those domains. This procedure also resulted in pieces of music with natural-sounding beginnings and endings.

Second, given that we chose to use improvisations as stimuli, we manipulated complexity by instructing the musicians who generated our stimuli to vary the complexity of their improvisations across five complexity levels. They were told to deliberately vary complexity across their improvisations while keeping the complexity within each improvisation relatively constant (for details of the procedure, see below). We chose to not formalize how the improvisers were supposed to vary complexity because such formalization might have endangered the ecological validity of the improvisations. This procedure relied on the judgments of those musicians who generated the stimuli, but provided a variation in complexity that was objective in the sense of being independent of the listeners in the study.

There are two types of independent variables used in this type of work: objective complexity and perceived complexity. This fact made our third methodological decision—choosing the independent variable for complexity—difficult because each type possesses desirable qualities. Objective complexity can be any manipulation in music. For example, using information theory in order to determine the selection, duration, and loudness of notes within a musical piece affords excellent control of stimulus complexity (see Vitz, 1966a). However, the use of information theory in this manner to explore the relationship between musical complexity and liking is limited because there is little similarity between the stimuli and actual music. Another way to use objective complexity is to vary melodic, harmonic, and rhythmic principles of music. For instance, Conley (1981) defined musical complexity as determined by the following musical dimensions: (a) regularity/irregularity of number of tones per chord, (b) number of independent parts, (c) number of different harmonies, (d) number of changes of harmony, (e) number of measures of tonic harmony, (f) number of measures of dominant harmony, (g) number of measures of nontonic/dominant harmony, (h) number of changes in rhythmic activity, (i) rate of rhythmic activity, and (j) duration. Others have used the same concept of varying musical dimensions to define objective musical complexity (Simon & Wohlwill, 1968; Smith & Melara, 1990). This process allows good control of stimulus complexity and results in stimuli with high ecological validity.

Perceived musical complexity is established by asking listeners to rate musical complexity after listening to a piece. Commonly, these ratings are collected by using Likert-like scales. Most often, a very brief or no description of what is meant by complexity is offered to the listeners before the ratings begin (Burke & Gridley, 1990; North & Hargreaves, 1995; Orr & Ohlsson, 2001; Radocy, 1982; Russell, 1982). This type of complexity measurement is desirable because it directly taps into how complex a piece of music sounds to the listener and can be used for any type of auditory stimulus.

For the present experiment, we choose perceived complexity as the independent variable because we take seriously the assumption that objective complexity measures are strongly correlated with perceived complexity measures (North & Hargreaves, 1995). Furthermore, with naturalistic stimuli such as ours, it would prove difficult to measure what objective factors contribute to complexity. The use of perceived complexity circumvents this problem.

Concerning our fourth methodological decision, we ensured that the participants in our study were true experts. The differences between novices, on the one hand, and music students and other subjects with limited expertise, on the other, might not be large enough to reveal behavioral differences with regards to the complexity/liking relationship. In fact, Smith and Melara (1990) found that the relationship between liking and transformation level (arguably a complexity manipulation) for undergraduate music majors was more similar to that for undergraduate nonmusic majors than that for graduate students in music and music professors. This suggests that the level of expertise is of crucial importance for the current work, as has been found in other domains of expertise (e.g., Gobet & Simon, 2000). In other words, a very high level of expertise should be incorporated to fully test the inverted-U hypothesis.

All of our expert participants had at least 10 years of musical training. This amount of training is consistent with how the expertise literature defines expertise (see Ericsson & Smith, 1991). Using years of training alone, however, is an inaccurate method of determining level of expertise. In fact, Ericsson, Krampe, and Tesch-Romer (1993) found that the number of cumulative practice hours is what differentiates very high levels of performance from good levels of performance, not the number of years of training. Therefore, we ensured that all of our participants had worked as professional musicians, as either teachers or performers, for at least 5 years. Professional musicians typically are performing at or close to the maximum performance levels acquired by musicians. In other words, our sample of musicians is likely to be among those with the most accumulated practice within a reasonable geographic area.

LINK TO PREVIOUS WORK

The methods and analyses used in the current study are identical to those used in a previous study (Orr & Ohlsson, 2001, Experiment 1) that involved 64 undergraduate psychology subject-pool participants. This study found support for the inverted-U hypothesis for bluegrass improvisations, but not for jazz improvisations. This finding suggested that the inverted-U relation might hold only for some musical styles. Although the cohort contained nonmusicians and moderately experienced musicians, Orr and Ohlsson did not use musical experience as a factor in the analyses. Therefore, in addition to collecting data on experts, the current study reintroduces the Orr and Ohlsson data in two ways. First, we split the undergraduate data set into two groups: nonmusicians and moderately trained musicians. By comparing the experts in the current study with nonmusicians and moderately trained musicians from the previous study, we will reveal how increasing expertise affects the complexity/liking relationship, albeit in a provisional manner. Second, we analyzed the full data set from the previous study (collapsing across the musical experience groups) in a novel way that has not previously been published. This method will be discussed in detail in the Methods and Results sections. Not all analyses in the current study included re-analysis of the undergraduate participants from Orr and Ohlsson.

Experiment

METHOD

Participants

Twelve professional jazz musicians and 10 professional bluegrass musicians participated as the expert listeners in this experiment. Their mean (*SD*) number of years of musical training was 28.4 (9.1) for the jazz musicians and 32.8 (9.6) for the bluegrass musicians. From Orr and Ohlsson (2001, Experiment 1), a no-training (n = 26) and a moderate training (n = 18) group were selected from the original 64 undergraduate participants by use of a musical experience index created from a musical experience questionnaire (this training analysis was not reported in Orr & Ohlsson, 2001). This index was based on the following five factors (*M* [*SD*] for the moderate-training group is in parentheses where appropriate; the no-training group had zero on the numerical factors and "no" on the categorical factors): initial age of musical training (7.7 [2.3] years), length of musical training (9.7 [4.1] years), whether musical training was presently ongoing (78 % yes), hours of practice per week (7.6 [5.4]), and whether musical training was ever undertaken (yes). These participants were drawn from the psychology subject pool at the University of Illinois at Chicago. In addition, for one of the analyses it was useful to include all 64 participants from Orr and Ohlsson (see the Regression Analysis Critique section below).

Materials

Two professional jazz musicians each created 20 short (30–40 s long) jazz improvisations and two professional bluegrass musicians each created 20 short bluegrass improvisations, for a total of 80 improvisations. These musicians did not participate as listeners. 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.

The musicians were instructed to make each improvisation match one of five complexity levels (*performer complexity*). Level 1—low complexity—was defined as "predictable, simple and uniform," whereas Level 5—high complexity—was defined as "unpredictable, surprising and erratic" (these terms were borrowed from North & 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-s pause after each improvisation. Similarly, the 40 bluegrass improvisations were placed in two random sequences, Bluegrass A and Bluegrass B.

Procedure

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 1 week apart. The sequences were counterbalanced across participants, so that approximately half the participants listened to either sequence (A or B) within each musical style. Each participant encountered the same two sequences, in the same order, in both sessions (e.g., Jazz B and Bluegrass B in both Sessions 1 and 2). As practice, participants rated liking or complexity for three improvisations at performer complexity Levels 1, 5, and 3 (in that order), without feedback, before each session.

Approximately half of the participants rated jazz improvisations in the first half of each session and bluegrass improvisations in the second half; the other half rated these styles in the opposite order. The participants were allowed a 3- to 5-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-s pause between successive improvisations. The listeners rated complexity on a seven-point scale where the anchors were equal to the performer complexity anchors. Liking was also rated on a seven-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." (See Appendix for detailed liking and complexity rating instructions.) The effects of the participants' assumptions about the relationship between complexity and liking (Sluckin, Hargreaves, & Colman, 1983) were controlled for by counterbalancing the order of the rating tasks. Approximately half of the participants rated complexity in their first session and liking in their second session; the other half performed these rating tasks in the opposite order.

At the end of the second session, the participants completed a musical background questionnaire.

Examples of Stimuli

Figures 2 through 7 present representative examples of the range of complexity produced by the improvisers. Figures 2, 3, and 4 represent the most simple, median, and most complex jazz improvisations, respectively, as rated by the expert musicians (collapsing across both

Fig. 2. The simplest jazz improvisation as rated by the jazz and bluegrass musicians. The mean complexity value was 2.5.

expert groups). Figures 5, 6, and 7 represent the same for the bluegrass improvisations. Because there was an even number of improvisations for each musical style, two improvisations served as candidate medians. We chose one of these at random for each musical style.

RESULTS

Complexity Judgments

To ensure that the performer complexity manipulation worked, we analyzed perceived complexity as a function of performer complexity. Perceived complexity should increase as performer complexity increases. Each analysis was conducted separately for the jazz and bluegrass improvisations and included both undergraduate groups (no-training and moderate training) and the two expert groups.

Fig. 3. The median (complexity) jazz improvisation as rated by the jazz and bluegrass musicians. The mean complexity value was 4.4.

Jazz Improvisations

A two-way (4 × 5) mixed analysis of variance (Expertise × Performer complexity) indicated that the four groups were not different in their overall perceived complexity ratings of the jazz excerpts, $F(3, 62) = 1.19$, *ns*,

Fig. 4. The most complex jazz improvisation as rated by the jazz and bluegrass musicians. The mean complexity value was 6.6.

MSE = 1.95. Collapsed across expertise, the increase in performer complexity was perceived, $F(4, 248) = 202.49$, $p < .001$, $MSE = 0.25$. These main effects were qualified by a reliable expertise by performer complexity interaction, $F(12, 248) = 3.71$, $p < .001$, $MSE = 0.25$. Figure 8 shows the mean perceived complexity by performer complexity for all groups.

Fig. 4. (Continued)

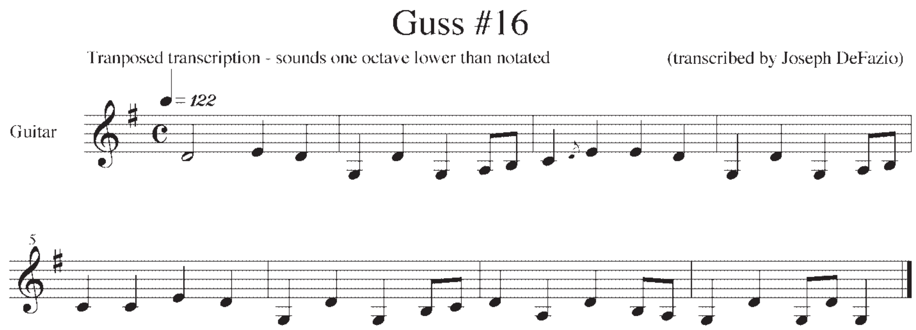

Fig. 5. The simplest bluegrass improvisation as rated by the jazz and bluegrass musicians. The mean complexity value was 1.1.

Bluegrass Improvisations

A two-way (4 × 5) mixed analysis of variance (expertise × performer complexity) indicated that the four groups were different in their overall perceived complexity ratings of the bluegrass excerpts, $F(3, 62) = 5.13$, $p < .01$, $MSE = 1.96$. Collapsed across expertise, the increase in performer

Fig. 6. The median (complexity) bluegrass improvisation as rated by the jazz and bluegrass musicians. The mean complexity value was 3.2.

complexity was perceived, $F(4, 248) = 161.55$, $p < .001$, $MSE = 0.29$. These main effects were qualified by a reliable expertise by performer complexity interaction, $F(12, 248) = 2.49$, $p < .01$, $MSE = 0.29$. Figure 8 shows the mean perceived complexity by performer complexity for each group.

Fig. 7. The most complex bluegrass improvisation as rated by the jazz and bluegrass musicians. The mean complexity value was 5.6.

It is clear that the performer complexity manipulation worked, both for the jazz and bluegrass improvisations. Follow-up analyses were not undertaken because comparisons between groups are not clearly interpretable because of the lack of a common anchor among the groups.

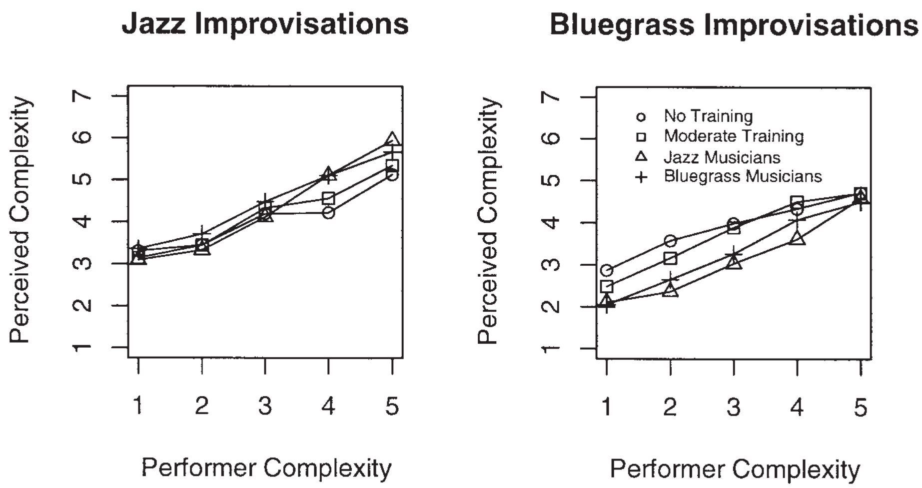

Fig. 8. Mean perceived complexity as a function of performer complexity for no-training undergraduates, moderate-training undergraduates, jazz musicians, and bluegrass musicians for both the jazz improvisations and bluegrass improvisations.

Liking as a Function of Complexity

In order to determine how musical training affects the relationship between complexity and liking, the mean liking rating for each improvisation was regressed onto the mean complexity rating for that improvisation. These analyses were performed separately for the jazz and bluegrass improvisations. In addition, the no-training and moderate-training groups from Orr and Ohlsson (2001) were included in these analyses.

We used the following two criteria to define an inverted-U shaped relationship between complexity and liking. First, the regression analysis must indicate that the quadratic component was a better fit statistically than was the linear component (in an inverted-U shape). The linear equation was interpreted as the correct interpretation when (a) it was statistically reliable, and (b) the quadratic equation did not explain more variance than did the linear equation, as tested by the statistical increment to R^2 test (Stimson, Carmines, & Zeller, 1981). The quadratic equation was interpreted as the correct interpretation when (a) it was statistically reliable, and (b) the quadratic equation explained more variance than the linear equation explained, as tested by the statistical increment to R^2 test. Notice that both the quadratic and linear equations might be statistically significant for the same data set. However, this does not imply that both interpretations are supported. The increment to R^2 test determines which one should be interpreted as correct (Stimson et al., 1981). For each figure that accompanies each regression analysis, the regression line represents the best-fitting line according to the criteria just listed. Second, for cases in which the quadratic component (inverted-U shape) was considered a better fit, the improvisations that exist above and below the apex of the equation will be separated into two groups. The correlation analyses between complexity and liking for each subset of improvisations (below and above the apex) will either support or reject an inverted-U relation between complexity and liking (henceforth, *apex-correlation analysis*). To support an inverted-U interpretation, the correlations should be positive to the left of the apex and negative to the right of the apex.

The validity of these regression analyses is questionable because of the low number of data points (40 improvisations for each musical style). Therefore, in addition to the regression and apex-correlation analyses described above, a subsampling procedure was implemented in which a randomly selected subsample of 37 improvisations served as the basis for both a regression and apex-correlation analysis. The subsampling procedure was conducted 100 times, with replacement, for each of the four groups. The logic for this subsampling is similar to that of the bootstrap (see Efron & Tibshirani, 1993 for the bootstrap) in that we are interested in getting an estimate of accuracy of the data. Typically, the bootstrap is used to get estimates of standard error; however, we are using subsampling as a way to test the accuracy of the shape and strength of the liking/complexity relation.

Interpretation of the resampling analyses is problematic because of the difficultly in justifying the removal of any of the data points. Each data point represents an individual improvisation, not an individual participant's score on a variable. We do not consider removing any of the improvisations as valid because they were developed in a controlled manner for the aims of this study. So, the resampling analyses are included to ease some readers' qualms about the low number of data points. However, we will use the full regression analyses (with 40 improvisations) as the bases for our claims.

Jazz Improvisations

The mean liking ratings for each improvisation were regressed onto the mean complexity ratings. The results are shown in Table 1 and Figure 9. Table 1 provides the R^2 and F values for both the linear and quadratic equations; the interpretation of the regression analyses (based on the criteria listed earlier); and the apex value and apex-correlation analysis for analyses in which the quadratic equation explained more variance than the linear equation explained.

Only the undergraduates indicated a reliable relation between liking and complexity. The no-training group showed a reliable quadratic relation. Only two data points were to the left of the apex, therefore nullifying the correlation analysis for these data points. A reliable negative rela-

TABLE 1
Summary of Liking/Complexity Regression Analyses

Group	R^2_{linear}	F_{linear}	R^2_{quad}	F_{quad}	Interpretation	Apex Value	Correlation Left Leg	Correlation Right Leg
Jazz Improvisations								
No training	.56*	47.79	**.66***	35.55	Quadratic	2.75	NI	-.83*
Training	**.28***	14.63	.33*	8.99	Linear	NR	NR	NR
Jazz musicians	.06	2.33	.15	3.31	No relation	NR	NR	NR
Bluegrass musicians	.04	1.48	.11	2.22	No relation	NR	NR	NR
Bluegrass Improvisations								
No training	.41*	26.42	**.51***	19.01	Quadratic	2.69	-.48	-.74*
Training	.01	0.30	**.27***	6.87	Quadratic	3.5	.55*	-.46*
Jazz musicians	.09	3.52	.13	2.67	No relation	NR	NR	NR
Bluegrass musicians	**.37***	22.19	.37*	11.00	Linear	NR	NR	NR

NOTE—Degrees of freedom for linear and quadratic regression analyses were (1,38) and (2,37), respectively.
*$p < .05$.
Boldface type highlights the R^2 of the interpreted line.
NI indicates not interpreted; NR indicates this cell is not relevant for the analysis.

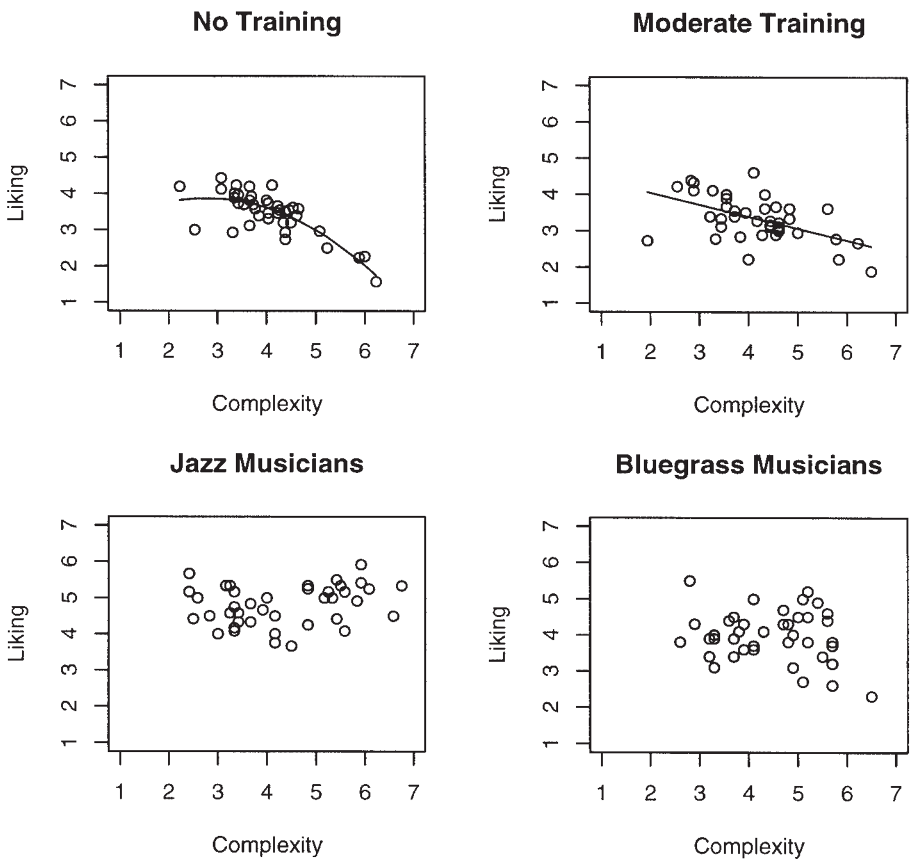

Fig. 9. Mean liking regressed onto mean complexity for the jazz improvisations for the no-training undergraduates, moderate-training undergraduates, jazz musicians, and bluegrass musicians.

tion existed among the 38 data points to the right of the apex. The moderate-training group exhibited a negative relation between liking and complexity.

Table 2 presents the results of the subsampling procedure and provides the following: *M* (*SD*) of the linear and quadratic R^2 values; the number of analyses interpreted as no relation, linear, and quadratic (quadratic was separated into two types; inverted U-shape and U-shape); *M* (*SD*) of apex values; the *M* (*SD*) correlation-coefficient for the left and right legs of the apex values, and the number of reliable correlations (see the asterisks in Table 2 for further explanation). Figure 10 presents the regression lines from the resampling analyses. For each group, the original 40 improvisations are plotted. However, each regression line represents one of the subsampling analyses. For subsampling analyses in which no relation was found between liking and complexity, no line was drawn.

It is clear from the subsampling analyses that the characterization of the data based on the full data set (all 40 jazz improvisations) was accurate, with the exception being the jazz musicians. Although the full data set revealed no relation between complexity and liking for the jazz musicians, about half of the subsampling analyses indicated that the quadratic u-shaped function was the best fit (average R^2_{quad} was .16), most of which indicated a reliable positive liking/complexity correlation for the right leg. This suggests that a weak relation might exist between liking and complexity for the jazz musicians.

Bluegrass Improvisations

The results for the bluegrass improvisations are shown in Table 1 and Figure 11. For both of the undergraduate groups, the quadratic equation

Table 2
Summary of Liking/Complexity Subsampling Analyses

			Interpretation					Correlation	
Group	R^2_{linear}	R^2_{quad}	**No Rel.**	**Linear**	**Invert U Quad**	**U Quad**	**Apex Value**	**Left Leg**	**Right Leg**
Jazz Improvisations									
No training	.56 (.05)	.66 (.03)	0	4	96	0	2.77 (0.28)	.90 (.19); 0	-.82(.02); 96*
Training	.28 (.06)	.33 (.05)	0	90	10	0	3.03 (0.13)	.91 (.13); 4*	-.55(.06); 10*
Jazz musicians	.06 (.02)	.16 (.03)	45	4	0	51	4.05 (0.07)	-.31 (.08); 4*	.54(.08); 47*
Bluegrass musicians	.04 (.02)	.11 (.03)	90	0	10	0	4.24 (0.11)	.07 (.16); 0	-.48(.04); 7*
Bluegrass Improvisations									
No training	.41 (.03)	.50 (.03)	0	0	100	0	2.65 (0.11)	-.41 (.21); 1*	-.72(.04); 100*
Training	.01 (.01)	.27 (.03)	0	0	100	0	3.46 (0.05)	.57 (.05); 92*	-.44(.05); 67*
Jazz musicians	.09 (.03)	.13 (.02)	78	22	0	0	NR	NR	NR
Bluegrass musicians	.37 (.03)	.38 (.03)	0	100	0	0	NR	NR	NR

Note—Numbers in parentheses are standard deviations.
*Number of statistically reliable correlations.

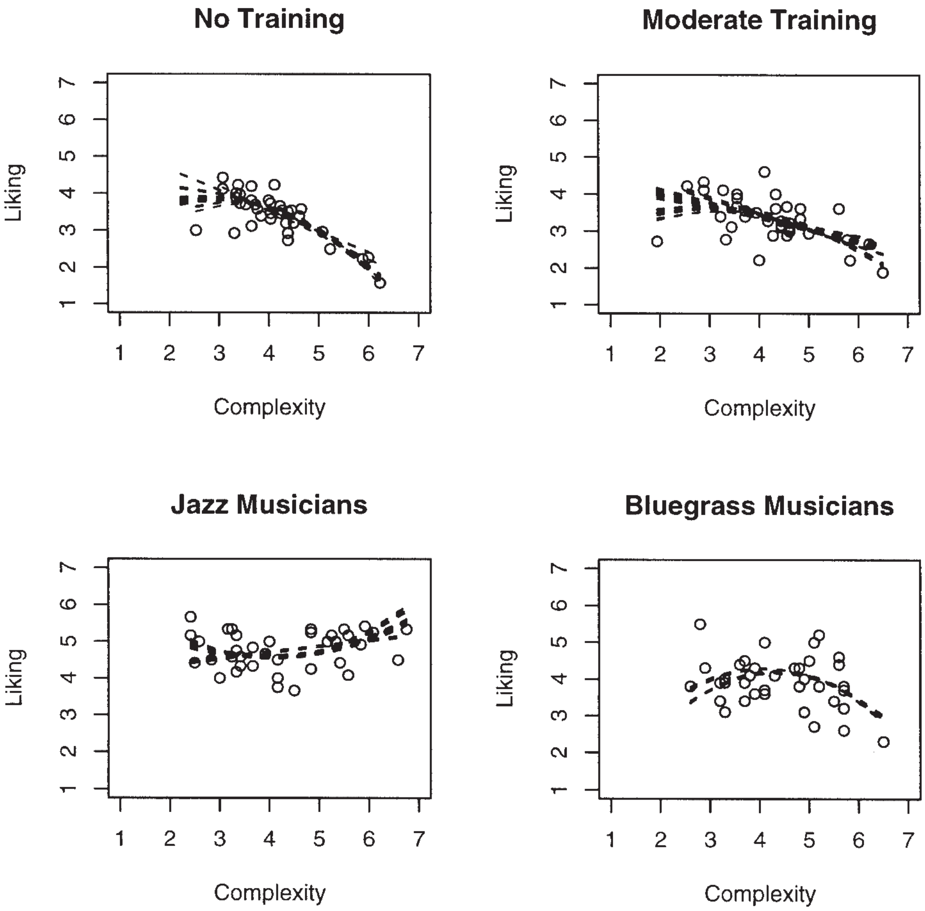

Fig. 10. Resampling analyses for the jazz improvisations. For each group, liking was regressed onto complexity 100 times; each regression used 37 randomly chosen improvisations.

was interpreted as the better-fitting equation. For the no-training group, the correlation between the 5 data points to the left of the apex was not reliable; the correlation was reliably negative for the 35 data points to the right of the apex. For the moderate training group, the correlation was reliably positive for the 17 data points to the left and reliable negative for the 23 data points to the right of the apex. The bluegrass musicians exhibited a linear relation between liking and complexity; for the jazz musicians, no relation was found.

The resampling analysis indicated that the full data set was an accurate characterization of the liking/complexity relationship, except for the jazz musicians (Figure 12, Table 2). About one fifth of the subsampling analyses showed a reliable linear relation between liking complexity for the jazz musicians.

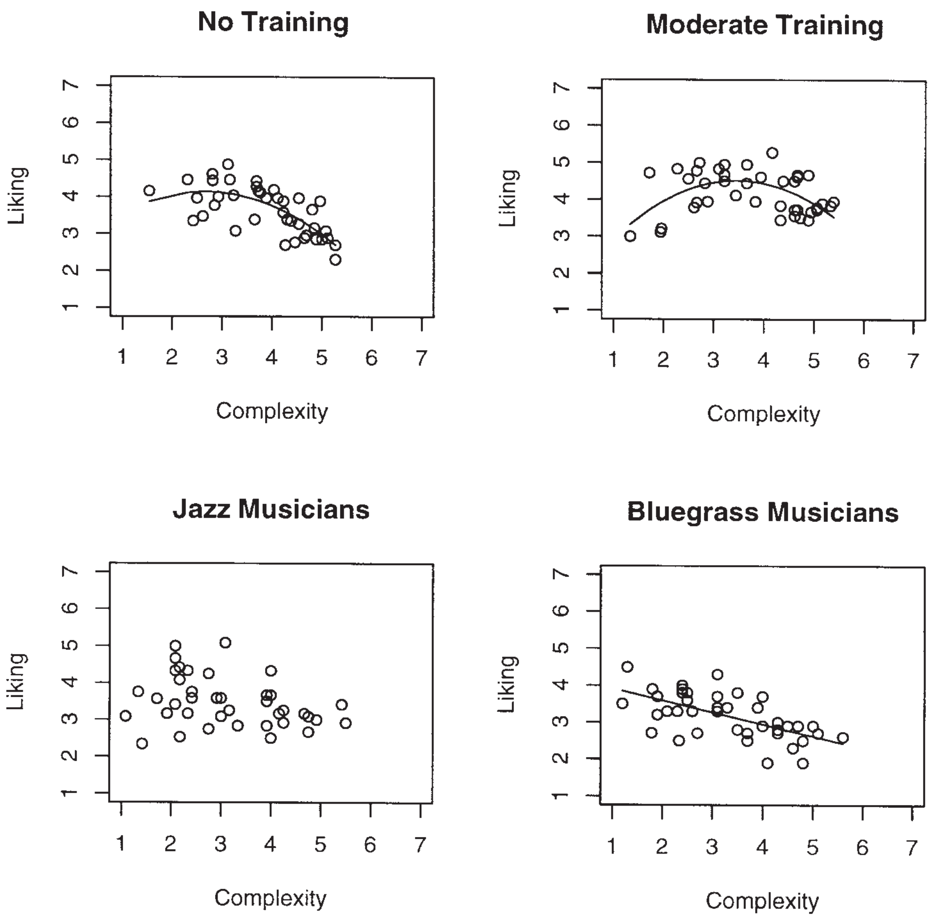

Fig. 11. Mean liking regressed onto mean complexity for the bluegrass improvisations for the no-training undergraduates, moderate-training undergraduates, jazz musicians, and bluegrass musicians.

Regression Analysis Critique

One criticism of the regression analyses is that the mean values for each improvisation may not be stable because of the low number of participants in each group (10 bluegrass musicians; 12 jazz musicians). If this was the case, then the nature of the complexity/liking relationships reported above might not reflect their true nature because of too much error per data point. This issue is important because the finding of no relationship between liking and complexity for the jazz improvisations for either the jazz or bluegrass musicians might be due to too much variation in each improvisation's mean value, in effect, washing out any relation between liking and complexity.

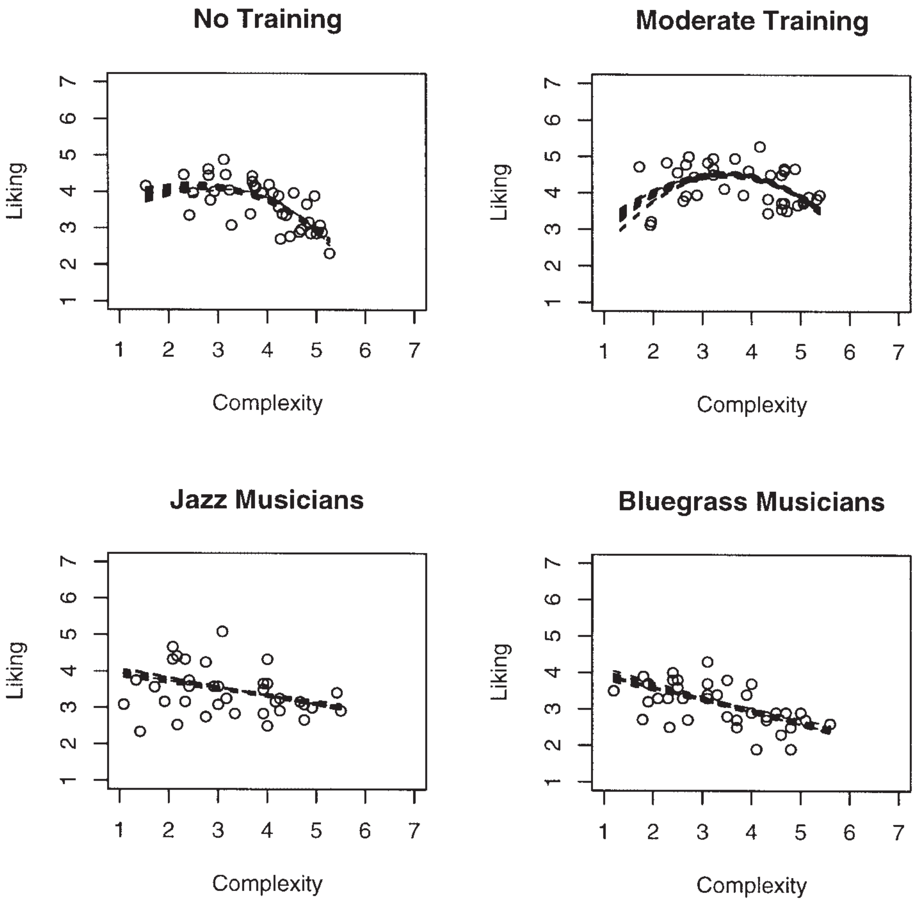

Fig. 12. Resampling analyses for the bluegrass improvisations. For each group, liking was regressed onto complexity 100 times; each regression used 37 randomly chosen improvisations.

To evaluate this potential problem, we computed a mean value that reflected the mean of the 80 improvisations' standard deviations for both the liking and complexity ratings. We did this separately for the jazz musicians, the bluegrass musicians, and the 64 undergraduates from the Orr and Ohlsson (2001, Experiment 1) study. If low *N* was a problem, the mean of the standard deviations should be reliably higher for the expert groups than for the undergraduate group because the expert groups had only 10 (bluegrass) and 12 (jazz) participants whereas the undergraduate group had 64 participants. This result was not found. For the complexity ratings, the means (*SD*) of the standard deviations for the undergraduates, jazz musicians, and bluegrass musicians were 1.35 (0.14), 1.26 (0.35), and 1.22 (0.28), respectively. These differences were reliable, $F(2, 237) = 4.34$, $p < .05$, $MSE = 0.07$. Furthermore, Tukey *a* procedure indicated that the only reliable difference was between the novices and bluegrass musicians ($p < .05$). For the liking ratings, the means (*SD*) of the standard deviations for the undergraduates, jazz musicians, and bluegrass musicians were 1.52 (0.14), 1.45 (0.24), and 1.52 (0.31), respectively. However, these differences were not reliable, $F(2, 237) = 2.06$, *ns*, $MSE = 0.06$. In short, the lower *N* for the jazz and bluegrass groups did not undermine the interpretation of the regression analyses.

Another criticism of the regression analyses is that the lack of a complexity/liking relationship for the jazz and bluegrass musicians listening to the jazz improvisations might be explained as resulting from half of the participants having a positive linear liking/complexity relationship and the other half having a negative linear relation liking/complexity relationship. If this were the case, the effect might be to obscure the liking/complexity relationship when data are collapsed across individuals. Only the expert musicians were included in this analysis because only they exhibited a null liking/complexity relationship. We tested this criticism by looking at each individual participant's complexity/liking regression analyses. Then we tallied how many of the participants exhibited each type of relation. Collapsing the data across the jazz and bluegrass participants, we found the following: for the jazz improvisations, three participants exhibited a negative linear relationship and two exhibited a positive linear relationship; for the bluegrass improvisations, seven exhibited a negative linear relation and one exhibited a positive linear relation. It appears, then, that for the jazz and bluegrass improvisations, the lack of relationship between complexity and liking is not due to equal positive and negative linear relations across individual participants.

Discussion

This experiment attempted to answer whether the inverted-U relation between liking and complexity holds for experts. We did not find any evidence that it does. Neither the jazz nor the bluegrass musicians exhibited an inverted-U relationship for either the jazz or bluegrass improvisations. In fact, the evidence for any relationship between complexity and liking was weak for the experts. For the jazz improvisations, liking was not related to complexity for either the jazz or bluegrass musicians. For the bluegrass improvisations, liking was only related to complexity for the bluegrass musicians and not for the jazz musicians.

This pattern of results raises two questions. First, why was there no relation between liking and complexity for either expert group listening to the jazz improvisations and for the jazz musicians listening to the bluegrass improvisations? Second, why was the liking/complexity relationship different between the jazz and bluegrass improvisations for the bluegrass musicians? These two questions will be dealt with in turn.

LACK OF LIKING/COMPLEXITY RELATION

One explanation for the lack of relation between liking and complexity for the experts might be due to a restriction of range in the complexity of the improvisations. This experiment used real music, created by professional musicians. Even the simplest and the most complex improvisations were real music; they did not consist of the same tone over and over again (very simple) or a random sequence of tones (very complex). Therefore, the full range of complexity of auditory stimuli extends below and above the level of the least and most complex improvisations provided in the current experiment. This pattern is clearly visible in the data: the lowest mean complexity of the jazz improvisations was above 2.4 for both the jazz and bluegrass musicians; for the bluegrass improvisations, the highest was below 5.7 (see Figures 9 and 11). Including simpler jazz improvisations and more complex bluegrass improvisations may have revealed a stronger liking/complexity relationship and/or an inverted-U relationship. However, a simple observation runs counter to this explanation. The range of perceived complexity was approximately equal among the expert and undergraduate groups (about four out of six points on the complexity scale). This range was large enough to reveal a relation between liking and complexity for the undergraduates; why should it not do the same for the experts. Furthermore, by including simpler and more complex stimuli, the ecological validity of the stimuli might have been compromised. Simply put, simpler jazz and more complex bluegrass improvisations might not be considered real music. The inverted-U hypothesis concerns music, not the full space of auditory events.

A second explanation for the lack of a liking/complexity relation suggests that complexity doesn't predict liking for experts because experts, when listening to music, are concerned with other aspects of music besides complexity. In support of this explanation, the literature suggests that some musical qualities are weighted more by experts than by nonmusicians during aesthetic judgments. For example, expressive characteristics are related to aesthetic judgments in music for experts (Clarke, 1993; Repp, 1992) but not for novices (Repp, 1992). Furthermore, is it clear that training as a musician involves the development of many aesthetic criteria such as timbre, time, harmony, and melody. These principles are applicable to any level of complexity. So, from a pedagogical perspective, complexity would not be considered a major aesthetic principle. As a first approximation, then, we subscribe to this explanation of these data.

Incorporating the two undergraduate groups allowed for a preliminary investigation into the nature of the function between expertise and the strength of the liking/complexity relationship. As is clear from Table 1, the amount of variation in liking explained by complexity (R^2) decreased as a function of expertise. Although the shape of this function is not well determined—only 3 data points sampled—the suggestion is that complexity becomes progressively less important for experts' liking judgments of music.

The current work served as an initial study into the relation between expertise, liking, and complexity that lacked the methodological problems of the past. Two directions for future research are apparent. First, to more accurately determine the shape of the function between expertise and the strength of the liking/complexity relation, a better sample should be used. Nonmusicians and several levels of expertise (e.g., 3, 6, 9, 12, 15 years of expertise) within one musical style should serve as an adequate sample to determine the nature of this function. Second, extensive use of verbal protocols (e.g., Ericsson & Simon, 1984) combined with on-line aesthetic data collection techniques (e.g., Madsen, Brittin, & Capperella-Sheldon, 1993; Orr & Miller, 1997) may provide insights into what musicians find likable in music and help to explain how aesthetic criteria change as a function of expertise.

BLUEGRASS MUSICIANS' BEHAVIOR DIFFERENT BY MUSICAL STYLE

The second question, why the bluegrass musicians exhibited a different type of liking/complexity relation for each musical style, is addressed next. At face value, this finding suggests that bluegrass musicians like simpler bluegrass improvisations but do not figure complexity into liking judgments when listening to jazz improvisations. High complexity alone cannot be responsible for the decrease in liking because the bluegrass musicians liked the most complex jazz improvisations. A speculation that falls in line with our claim that musical training develops specific aesthetic criteria is that the bluegrass musicians dislike the more complex bluegrass improvisations because of a learned aesthetic value system that disdains complexity in bluegrass. A peculiarity of this interpretation is that it assumes that the bluegrass listeners apply this aesthetic criterion (e.g., the hypothesized disdain for complexity) only to the bluegrass improvisations and not to the jazz improvisations.

Before making solid conclusions based on the bluegrass improvisations, it seems necessary to replicate this study with a new set of improvisations created by different improvisers and to use verbal protocol analysis. Until such data is forthcoming, we do not make strong claims on this issue.

IMPLICATIONS

These data suggest that increased musical expertise reduces the importance of complexity in aesthetic judgments of music. This is apparent even when comparing only the no-training and moderate-training undergraduates. Furthermore, despite the difficulty in interpreting the bluegrass musicians' data for the bluegrass improvisations, it remains clear that the liking/complexity relation was weaker for the bluegrass musicians than for the no-training musicians. So, the general principle of an inverse relation between the strength of the liking/complexity and expertise is founded in the data.

The use of two musical styles proved to be very informative. In doing so, it is clear that the complexity/liking relation might be different for different musical styles. This difference was apparent in the undergraduate data (the inverted-U held only for bluegrass but not for jazz) and the expert data (complexity was not a factor at all for the jazz musicians and was a factor only for bluegrass for the bluegrass musicians). These results imply that aesthetic behavior in music may change as a function of musical style in terms of both the musical stimuli and the participants listening to the music. The overarching principle suggested from these data, although provisional at this point, is that expertise includes the learning of aesthetic criteria. In the absence of learned aesthetic criteria, complexity more strongly predicts liking for music, unless complexity itself serves as one of the learned criteria.

References

Abeles, H. F., & Chung, J. W. (1996). Response to music. In D. A. Hodges (Ed.), *Handbook of music psychology* (2nd ed., pp. 285–342). San Antonio, TX: Institute for Music Research.

Arnheim, R. (1986). The other Gustav Theodor Fechner. In S. Koch & D. E. Leary (Eds.), *A century of psychology as a science* (pp. 855–865). New York: McGraw-Hill.

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

Berlyne, D. E. (Ed.). (1974). *Studies in the new experimental aesthetics.* Washington, DC: Hemisphere.

Birkhoff, G. D. (1933). *Aesthetic measure.* Cambridge, MA: Harvard University Press.

Burke, M. J., & Gridley, M. C. (1990). Musical preferences as a function of stimulus complexity and listeners' sophistication. *Perceptual and Motor Skills, 71,* 687–690.

Chandler, A. R. (1934). *Beauty and human nature.* New York: D. Appleton-Century Company.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The nature of expertise.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Clarke, E. F. (1993). Imitating and evaluating real and transformed musical performances. *Music Perception, 10,* 317–341.

Conley, J. K. (1981). Physical correlates of the judged complexity of music by subjects differing in musical background. *British Journal of Psychology, 72,* 451–464.

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, DC: Hemisphere.

Davies, J. B. (1978). *The psychology of music.* London: Hutchinson.

Davis, R. C. (1936). An evaluation and test of Birkhoff's aesthetic measure formula. *Journal of General Psychology, 15,* 231–240.

Day, H. Y. (1967). Evaluations of subjective complexity, pleasingness and interestingness for a series of random polygons varying in complexity. *Perception and Psychophysics, 2,* 281–286.

Dember, W. N., & Earl, R. W. (1957). Analysis of exploratory, manipulatory, and curiosity behaviors. *Psychological Review, 64,* 91–96.

Dorfman, D. D. (1965). Esthetic preference as a function of pattern information. *Psychonomic Science, 3,* 85–86.

Dorfman, D. D., & McKenna, H. (1966). Pattern preference as a function of pattern uncertainty. *Canadian Journal of Psychology, 20,* 123–153.

Eckblad, G. (1963). The attractiveness of uncertainty. *Scandinavian Journal of Psychology, 4,* 1–13.

Eckblad, G. (1964). The attractiveness of uncertainty: II. The effects of different rates of reduction in the level of subjective uncertainty. *Scandinavian Journal of Psychology, 5,* 33–49.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Ericsson, K. A., Krampe, R. T., & Tesch-Romer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100,* 363–406.

Ericsson, K. A. & Simon, H. A. (1984). *Protocol analysis: verbal reports as data.* Cambridge, MA: MIT Press.

Ericsson, K. A., & Smith, J. (Eds.). (1991). *Toward a general theory of expertise: Prospects and limits.* New York: Cambridge University Press.

Eysenck, H. J. (1941). The empirical determination of an aesthetic formula. *Psychological Review, 48,* 83–92.

Eysenck, H. J. (1942). The experimental study of the 'Good Gestalt': A new approach. *Psychological Review, 49,* 344–364.

Eysenck, H. J. (1968). An experimental study of aesthetic preference for polygonal figures. *The Journal of General Psychology, 79,* 3–17.

Gabrielsson, A. (2001). Emotions in strong experiences with music. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotions: Theory and research* (pp. 431–449). Oxford, UK: Oxford University Press.

Gobet, F., & Simon, H. A. (2000). Five seconds or sixty? Presentation time in expert memory. *Cognitive Science, 24,* 651–682.

Guilford, J. P. (1934). The affective value of color as a function of hue, tint and chroma. *Journal of Experimental Psychology, 17,* 342–370.

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

Kamman, R. (1966). Verbal complexity and preference in poetry. *Journal of Verbal Learning and Verbal Behavior, 5,* 536–540.

Madsen, C. K., & Brittin, R. V., & Capperella-Sheldon, D. A. (1993). An empirical method for measuring the aesthetic response to music. *Journal of Research in Music Education, 41,* 57–69.

Martin, L. J. (1906). An experimental study of Fechner's principles of aesthetics. *Psychological Review, 13,* 142–219.

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 & R. Neperud (Eds.), *The foundation of aesthetics, art, and art education* (pp. 7–42). New York: Praeger.

McMullen, P. T. (1996). The musical experience and affective/aesthetic responses: A theoretical framework for empirical research. In D. A. Hodges (Ed.), *Handbook for music psychology* (2nd ed., pp. 387–400). San Antonio, TX: Institute for Music Research.

McMullen, P. T., & Arnold, M. J. (1976). Preference and interest as functions of distributional reduandancy in rhythmic sequences. *Journal of Research in Music Eduaction, 24,* 22–31.

Moreland, R. L., & Zajonc, R. B. (1976). A strong test of exposure effects. *Journal of Experimental Social Psychology, 12,* 170–179.

Munsinger, H., & Kessen, W. (1964). Uncertainty, structure, and preference. *Psychological Monographs, 78,* 1–24.

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

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

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

Orr, M. G., & Miller, L. K. (1997, August). *Continuous measurement of the aesthetic response to music: Genre and training effects.* Paper presented at the Semi-Annual Conference of the Society for Music Perception and Cognition, Cambridge, MA.

Orr, M. G., & Ohlsson, S. (2001). The relationship between musical complexity and liking in jazz and bluegrass. *Psychology of Music, 29,* 108–127.

Pierce, E. (1894). The aesthetics of simple forms. *Psychological Review, 1,* 483–495.

Pierce, E. (1896). The aesthetics of simple forms. *Psychological Review, 3,* 270–282.

Radocy, R. E. (1982). Preference for classical music: A test for the hedgehog. *Psychology of Music, Special Issue,* 91–95.

Repp, B. H. (1992). A constraint on the expressive timing of a melodic gesture: Evidence from performance and aesthetic judgment. *Music Perception, 10,* 221–242.

Rump, E. E. (1968). Relative preference as a function of the number of elements in abstract designs. *Australian Journal of Psychology, 27,* 39–48.

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

Scherer, K. R., & Zentner, M. R. (2001). Emotional effects of music: Production rules. In P. N. Juslin & J. A. Sloboda (Eds.), *Music and emotion: Theory and research* (pp. 361–392). Oxford, UK: Oxford University Press.

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

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

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

Stimson, J. A., Carmines, E. G., & Zeller, R. A. (1981). Interpreting polynomial regression. In P. V. Marsden (Ed.), *Linear models in social research* (pp. 88–96): Beverly Hills, CA: Sage.

Vitz, P. C. (1964). Preference for rates of information presented by sequences of tones. *Journal of Experimental Psychology, 68,* 176–183.

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

Vitz, P. C. (1966b). Preference for different amounts of visual complexity. *Behavioral Science, 11,* 105–114.

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

Washburn, M. F. (1911). A note on the affective value of colors. *American Journal of Psychology, 22,* 114–115.

Wohlwill, J. F. (1968). Amount of stimulus exploration and preference as differential functions of stimulus complexity. *Perception and Psychophysics, 4,* 307–312.

Zajonc, R. B. (1968). Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology, 9*(Monograph Suppl. 2, Pt. 2).

Appendix: Instructions for the Complexity and Liking Ratings

COMPLEXITY RATING TASK

Over the next 50 minutes you will hear 80 pieces of music. Each will last for 30 seconds. You will be rating how "complex" you think these pieces are. "Complex" means how easy it is to predict what the music will do next and how many surprises the music contains. More complex pieces are harder to predict. At the end of each excerpt, you should rate the extent to which you think that a piece is complex. You will have 10 seconds to give your rating. Try to make these ratings independently of the level of complexity in the music you normally listen to. There are no right or wrong answers—your honest opinion is what counts. Try to use the full range of the rating scales (i.e., do not be afraid to give ratings of 1 or 7).

Please give your complexity ratings for the following pieces on a scale of 1 to 7, where 1 = "very low complexity" (i.e., very predictable, simple, and uniform), 7 = "very high complexity" (i.e., very unpredictable, surprise, and erratic), and 4 = midway between the two.

Before we begin, please rate the following three practice pieces.

	Low Complexity			Midway			High Complexity
Practice	1	2	3	4	5	6	7

LIKING RATING TASK

Over the next 50 minutes you will hear 80 pieces of music. Each will last for 30 seconds. You will be rating how much you like these pieces. At the end of each excerpt, you should rate the extent to which you liked it. You will have 10 seconds to give your rating. Try to rate your liking for the segments independently of your liking for the music you normally listen to. There are no right or wrong answers—your honest opinion is what counts. Try to use the full range of the rating scales (i.e., do not be afraid to give ratings of 1 or 7).

Please give your liking rating for the following pieces on a scale of 1 to 7 where 1 = "did not like at all," 7 = "liked very much," and 4 = midway between the two.

Before we begin, please rate the following three practice pieces.

	Did Not Like at All			Midway			Liked Very Much
Practice	1	2	3	4	5	6	7