The Effects of Literal Phrase Repetition on Listener Preferences in Popular Music

Questions about the nature of musical form have some of the most elusive answers in music theory today. One basic problem lies in the inability to integrate two conflicting definitions of the word "form" into a single concept. These two definitions contrast the "outer" form of a work – i.e. what elements a work shares with others – with the "inner" form of a work – i.e. those unique qualities that make a piece different from any other. Bonds (1991) describes this opposition as a paradox between the "conformational" approach (outer form) and the "generative" approach (inner form). Bonds goes on to say: "There is a disturbing absence of any theoretical basis of form that can reconcile the generative and conformational approaches in a convincing fashion" (p. 29).

A good example of the difficulty in circumscribing exactly what "form" is can be found in Burnham (2002). In this essay, Burnham gives a history of form in music theory. The sonata is Burnham's connective thread, by which he explores the tension in sonata theory between the harmonic view (a conformation approach) and the thematic view (a generative approach). Yet for all the unresolved issues with respect to sonata form, one wonders how a resolution might help explain other musical forms or styles. For example, with the dissolution of harmony and theme in much 20th-century music, an answer to the dilemma of sonata form does not promise to address basic features of many other structures. As a result, can we have a unified definition of form that responds equally well to a broader scope of musical styles?

This turbid state theory on musical form becomes even more opaque when investigating popular music. Popular music theory – a relatively new discipline – inherits both the strengths and weaknesses of classical music theory. Covach (2005), for instance, takes the traditional tact that harmony tends to be the main factor in delineating formal units. The traditional definition of phrase – if applied to popular music – also depends heavily on harmonic motion (Laitz, 2003). Yet these harmonic approaches provide no help when faced with genres of popular music (rap, metal, blues, etc.) in which both large-scale and local harmonic motion is often static. How can we understand phrases and formal structure in such settings without better theoretical tools? It is in attempt towards an answer that cognition offers hope since phrases and formal structures are often so perceptually clear.

One reoccurring trend in the writings of composers on musical process is the balance between variety and unity (e.g. Stravinsky, 1942). Often this interplay between variety and unity is recast in terms of variation and repetition. Different authors invoke different terminology for these concepts (parallelism, similarity, etc.), but the basic premise seems universal: repetition endows a musical work with cohesiveness, while variation imparts interest. Harmonic change/consistency is only a single factor in this balance. Other domains – melody, texture, timbre, etc. – all contribute to the overall effect. In response to this challenge, theorists including Meyer (1989) and Ockelford (1991) have developed systems that directly investigate the many roles of variety and unity in music, but these systems are still operative on more of a metatheoretical than practical level.

The nagging questions of musical form still remain. How, therefore, might the field of music cognition help us better understand form? Specifically, why do compositions take on any given form over another? How necessary is the particular form of a composition to its appreciation? From a perceptual perspective, what effect (if any) do changes in form have on the listener? Through the investigation of such questions, we may hope to gain a deeper knowledge of the nature of musical structure.

BACKGROUND

Prior Research:

Only a handful of experiments have looked at what sort of effects the structural alteration of a composition's form has on the listener. In general, most of these studies have had difficulty in showing much if any change in response from participants. These structural alterations have involved a few different methods, which range from rearrangements of a single work to more dramatic alterations of multiple pieces.

One of the earliest studies into the effects of structural rearrangement was conducted by Gotlieb and Konečni (1985). In this study, a recording of J. S. Bach's *Goldberg Variations* was used to create two altered versions, each of which included a different ordering of the variations from the original sequence. Participants rated one of the versions (original or altered) on 15 scales, including ugly/beautiful, emotional/unemotional, simple/complex, etc. The authors found no consistent differences in the ratings between the altered versions and the original, however, which thereby implies that structural rearrangement of the composition had very little effect on subjects. Although one effect was noticed (the original-order version was rated as significantly "warmer"), the authors' p-value for significance was set at 0.05, and thus it is not statistically surprising that one out of the 15 categories showed some sort of effect.

Gotlieb and Konečni's experiment came under sharp criticism by Batt (1987). Batt argued that variation-form compositions withstood structural rearrangements much more robustly than other formal types due to the harmonically self-contained nature of the each variation. Batt argued that rearrangements of other formal types, such as sonata form, would create more drastic effects on listeners due to sonata form's interwoven thematic character. To provide a counterexample, Batt posited that the first movement of Mozart's 40th symphony (K. 550) in G minor would show clear effects of structural rearrangements on listeners.

Karno and Konečni answered Batt's challenge in their 1992 article. The Mozart G minor symphony movement was used to create four different altered versions, which differed in the sequence of major sections (first theme, transition, second theme, etc.). Participants listened to all five versions (original plus altered versions) in one listening session and rated each on three scales: preference, desire to own, and interestingness. Similar to the results of the 1985 study, though, the authors found no significant effects between the original and altered versions. What the authors did find (although somewhat downplayed) was that listeners tended to rate highest whichever version was heard first. Since this first movement has a relatively long duration (~ 8 minutes) and is saturated in thematic events, it may not be surprising that further repeated listenings created decreases in preference ratings. A fatigue effect (as order effect) thus apparently influenced results. The authors may have shown that preferences for the original version could not overcome this fatigue effect, therefore, but they could not conclusively prove that listeners' preferences were immune to the effects of structural rearrangements.

The effects of smaller-scale structural alterations were explored by Tillman and Bigand (1996). In this study, three compositions – one each by J. S. Bach, W. A. Mozart, and Schoenberg – were divided into 4-bar phrases. The authors created an altered version of each in which the order of these 4-bar chunks was reversed from the original piece, e.g. the last four bars became the first four, the penultimate four bars became the second four, etc. Participants rated the "forwards" and "backwards" versions on 27 scales of expression. No significant effects of the alteration were found, however. Each composers' work – whether original or altered – had its own expressivity profile as distinct from the others.

Partially due to this seemingly consistent lack of experimental proof for any listener effects with regard to structural rearrangement, Tan and Spackman (2005) took a more radical approach in their study. Instead of rearranging a single musical work, the authors created composite compositions ("patchworks") that sequentially merged 20-second excerpts from three different solo piano pieces. As well, the authors created repeated compositions, in which a single 20-second excerpt from a solo piano piece was repeated three times. Participants rated both the altered version as well as unaltered solos on the basis of perceived "unity." The results clearly showed that listeners perceived the patchwork compositions to be significantly less unified than the unaltered or repeated compositions. More generally speaking, the authors were able to show that certain structural alterations did produce effects on listeners – effects that had eluded prior experimenters. This study was still unable to find significant differences in structural alterations within the same composition, however.

In Tan, Spackman, and Peaslee (2006), the authors followed-up on their prior study to investigate the effects of multiple hearings on listener preferences. Over the course of two days, participants heard a total of four hearings for each of the patchwork and unaltered compositions. In general, preference ratings for the patchwork compositions increased while the preference ratings for the unaltered compositions declined. By the fourth hearing, in fact, preference ratings of the patchwork compositions actually exceeded those of the unaltered compositions. In other words, listeners – through the effect of repetition alone – began to prefer randomly-organized compositions to those pieces in which the composers' original design was left intact. How then does this perhaps surprising result inform our notion of form and its function in musical perception? Presuming that a piece's particular form does matter (and is somehow superior to random organization), what explanation can be given for such a finding?

Aesthetic Theory:

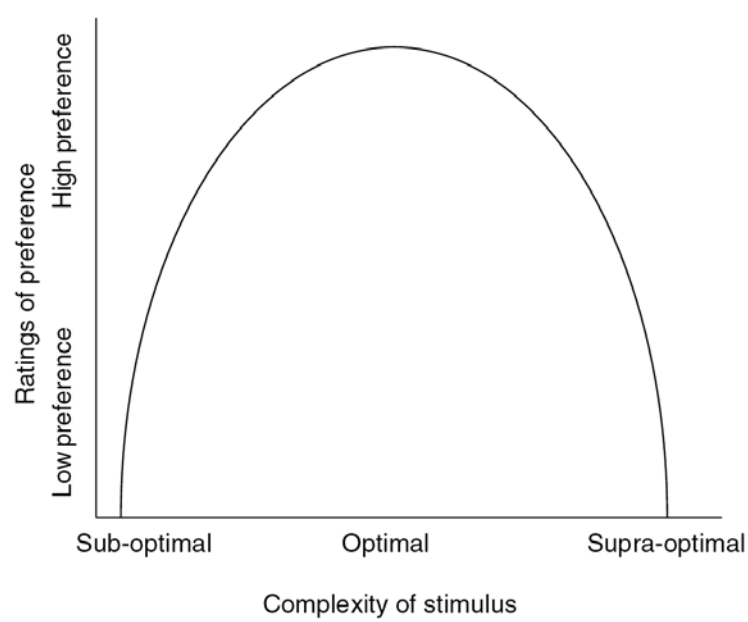
The authors of Tan et al. (2006) discuss their results with respect to theories from the field of aesthetics. As test cases, they present two competing theories that might predict the results of their experiment on repeated listenings. The first theory, known as the mere exposure theory, is drawn from Zajonc (1968); the other, known as the two-factor arousal theory, is based on the work of Berlyne (1971).

Zajonc's mere exposure hypothesis predicts that exposure to a stimulus is all that is required to "enhance" a subject's "attitude" towards this stimulus. For example, the more frequently a participant would hear an excerpt, the higher the participant's preference rating would be for this excerpt. In the case of the Tan et al. 2006 study, this mere exposure hypothesis seems contradicted by the decrease in preference ratings for unaltered compositions over multiple hearings. Preferences ratings for the unaltered compositions should have increased, according to Zajonc's hypothesis, not decreased. Such an exposure effect does appear to be operative with the patchwork stimuli, however, where preferences did increase with exposure.

Berlyne's two-factor arousal theory offers an alternate prediction to Zajonc's. Berlyne's theory is based on the fundamental premise that a subject's preference for a stimulus is highest when the stimulus is engendering a moderate level of arousal in the subject. "Arousal" – as defined by Berlyne – is a subjective factor, however, and the level of arousal for any given stimulus necessarily differs from subject to subject. The reason for the relative nature of arousal is tied to the components that contribute to this state of arousal. Certain properties – such as complexity and novelty for example – increase arousal, while others – such as simplicity and familiarity – decrease arousal. Inherently, these factors vary from subject to subject and thus arousal levels are subject-dependent. The level of novelty/familiarity with a particular stimulus will rarely be equivalent between two people. In a similar way, the level of perceived complexity for a given stimulus will vary from subject to subject: highly-trained jazz musicians, for instance, may perceive a particular jazz solo as only moderately complex, whereas an infrequent consumer of jazz might perceive the same solo as fairly complex.

In essence, therefore, Berlyne's theory predicts – all other parameters being ideal – that preferences will be highest when a single factor is at its optimal level. As this factor moves away from its optimal level – whether via an increase or decrease – preferences will necessarily decrease. The relationship between these two factors is typically represented by an inverted-U (i.e. quadratic) curve. Figure 1 shows this relationship using complexity as an example:

Figure 1: The inverted-U relationship between complexity and preference (Berlyne)



Experimental support for the inverted-U relationship between subjective complexity and preference can be found in numerous studies (e.g. Hargreaves, 1984; Orr and Ohlsson, 2005). In Tan et al. 2006 as well, the authors posited that their results were due to this inverted-U relationship. The authors argue that on the initial hearing, the relatively high subjective complexity and high novelty for the patchwork compositions led to high levels of arousal and thus lower-than-peak preferences. As the patchwork compositions were heard multiple times, novelty decreased, and thus the overall arousal level for these stimuli – which was initially at a higher-than optimal level – decreased such that preferences actually increased. We may imagine a shift to the left on the inverted-U curve with a decrease in arousal. Such a shift engenders an increase in preference on the right-hand side of the curve (where the slope is negative) and a decrease in preference on the left-hand side of the curve (where the slope is positive).

One final thought worth mentioning is that the theories of Zajonc and Berlyne are not mutually exclusive, despite their seemingly incompatible hypotheses. It is important to note that Zajonc's hypothesis appears to have been developed to account for effects with unfamiliar stimuli. As such, the novelty of these stimuli would be high and thus correspond to a high level of arousal and low level of preference. Zajonc, therefore, may be seen as describing the situation as captured by the right-hand side of Berlyne's inverted-U relationship, where increased familiarity correlates to an increase in preference. Experimental support for Zajonc's theory might thus be investigating only one aspect of Berlyne's broader hypothesis.

GOALS

Experimental work involving structural alterations of musical compositions has therefore mainly focused on a single overarching approach: rearrangement of musical segments, either within a single piece or between multiple pieces. Although Tan and Spackman (2005) used repetitive stimuli in their study, these stimuli were not a primary concern of their experimental design or results. Very little work has thus been devoted to investigating the effects of structural repetition within a work. Since repetition – along with variation – is so central in composers' descriptions of music, the function of structural repetition begs closer attention. Repetition, of course, may occur in multiple domains. Harmonies may repeat (e.g. chaconne), melodies may repeat (e.g. ostinato), instrumentations may repeat (e.g. solo suite), etc. For the purposes of this study, repetition is explored through the literal repeat of an entire musical phrase (e.g. four bars). All domains are thus involved in the process of repetition.

Popular music seems like an ideal source of musical material to investigate the role of such repetition since literal phrase repetition occurs in many different settings. Choruses, for example, typically undergo multiple presentations during the course of a song. Moreover, these choruses commonly appear with the exact same instrumentation, melody, harmony, etc. Are listeners sensitive to the number of times these choruses are repeated, and if so, to what effect?

Hypotheses on the effect of literal phrase repetition on listeners can be grounded in the aesthetic theories discussed earlier. Following Berlyne's work, for example, phrase repetition should create a decrease in novelty for listeners and thus a decrease in arousal. This decrease in arousal can be mostly directly measured in terms of preference ratings, since preferences are directly tied to arousal. The effect of a decrease in arousal on preference, however, depends on the overall level of the listener's arousal, which may be higher or lower than optimal. In order to gauge the level of a listener's arousal, therefore, an alternate measure needs to be used aside from novelty or preference. In this study, subjective complexity serves as this rough barometer. If complexity and preference ratings are given for a stimulus, a general map of arousal level can be created against which the effect of changes to this stimulus can be related.

In summary, we expect (and Berlyne's theory predicts) conformance to the effects of the inverted-U relationship between arousal and preference. Specifically, we expect that stimuli for which complexity ratings are high and preferences ratings are low should encounter an increase in preference ratings due to repetition. Conversely, stimuli for which complexity ratings are low and preference ratings are also low should encounter a decrease in preference ratings due to repetition.

METHODS

Participants:

15 subjects (9 males, 6 females) participated in the pilot study. All subjects were in their 20s or 30s and listened to popular music. The level of musicianship varied greatly among subjects and was somewhat difficult to measure, especially since some subjects had played instruments (guitar, bass, etc.) for many years but had never received much if any classroom instruction. In order to make a clear distinction, therefore, subjects were also categorized by formal training, which was defined as having earned a college-level degree or higher in music. Three subjects had formal training and 12 did not. Additionally, it should be noted that three of the subjects were familiar with the goals of the study (a potentially-biased group), while 12 subjects were blind (an unbiased group). Two of the three subjects in the potentially-biased groups were also two of the three formally-trained musicians. The inability to clearly separate these two factors (bias and training) presents a challenge that will be discussed later.

Materials:

A total of 72 stimuli were used in this study. These stimuli were created from 18 different songs drawn from a wide variety of popular styles and time periods (see Appendix). The main criterion in determining which songs would be used was the ability of the chorus for each song to be repeated without causing any obviously poor transitions or glaringly problematic structural changes. In an effort to keep these structural changes fairly consistent from song to song, a further criterion was added that these chorus sections had to be 4 measures of 4/4 time.

18 stimuli consisted of just a single iteration of the chorus alone. These chorus sections ranged in duration from 7 seconds to 14 seconds. The remaining stimuli were longer excerpts that were the result of creating three versions of each of the 18 songs, where each version differed by the number of times the chorus appeared. These three versions represented: A) no repetition (a single presentation of the chorus), B) repetition (two presentations of the chorus), and C) high repetition (four presentations of the chorus).

In order to keep the duration of the three versions of each song equal, non-chorus material from the same song was substituted as filler for chorus material when needed. The total length of these longer versions equaled five times the length of the chorus and therefore ranged in duration from about 35 seconds to 1 minute 10 seconds. The choruses were centered within these three versions, as Table 1 below shows. Non-chorus material was chosen on its viability to contrast the chorus; this non-chorus material included pre-choruses, verses, bridges, instrumental breaks, etc., none of which ever included literal phrase repetition (despite the generic label "A" used below). Often, there was no "original" version of the song used in this study since the creation of these variable-repetition-level versions necessitated customized song structures.

Table 1: Structural arrangement of longer stimuli based on repetition level

Repetition level	**Number of chorus presentations**	**Structure of stimulus version (A = non-chorus material)**
no repetition	1x	A – A – CH – A – A
repetition	2x	A/2 – A – CH – CH – A – A/2
high repetition	4x	A/2 – CH – CH – CH – CH – A/2

Procedure:

To minimize the effects of repeated hearings, participants were asked to submit to multiple listening sessions – each on a different day – in which unique versions of each of the 18 songs would be heard. To help facilitate the participation of subjects over multiple days, a web interface was created that allowed subjects to login and complete each progressive step of the study according to their own schedules. In the first listening session, subjects heard the chorus-only excerpts of each song. Participants were asked to rate this standalone chorus in terms of both preference and perceived complexity on scales of 1-9. Subjects were encouraged to use the entire range of these scales. The order of choruses was randomized for each subject.

On listening sessions 2, 3, and 4, participants were presented with the longer-version stimuli in which chorus iterations were manipulated. Participants then rated these longer versions on preference scales of 1-9. Subjects heard an equal number of 1x, 2x, and 4x versions on each session, but which version of which song appeared in what order on which session was randomized. In each listening session, therefore, each subject would hear 18 long excerpts – one version from each song – but no duplication as to the sequence of these versions occurred between subjects. All subjects eventually heard and rated all versions of all songs a single time.

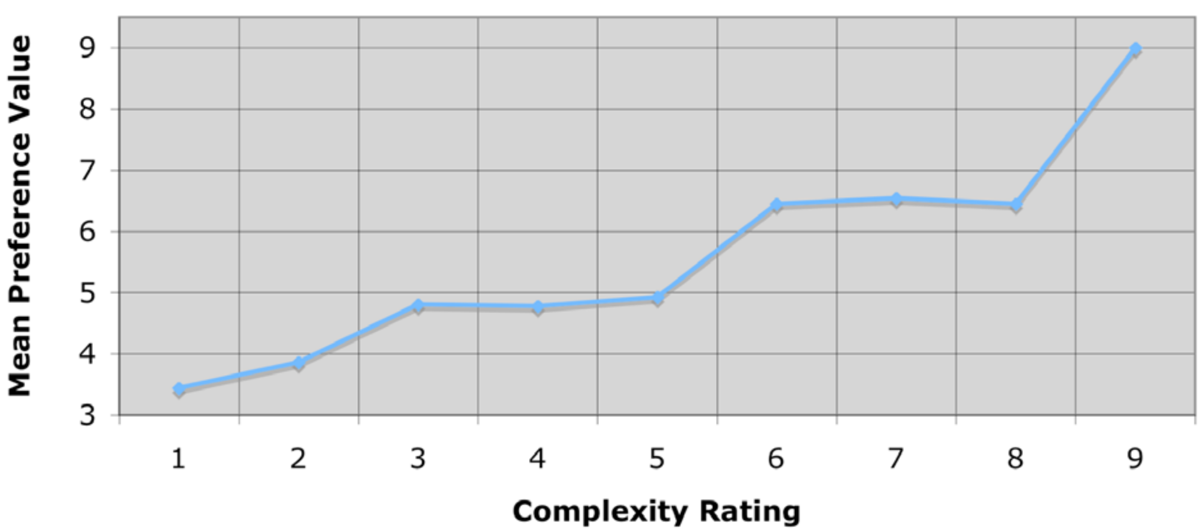
In addition to the advantages of complete randomization and a flexible completion schedule, the use of a web application inherently had a number of disadvantages. The foremost disadvantage was that the listening environment could not be controlled. Subjects, for example, could potentially not listen to the entire duration of each stimulus, which would compromise any intended effects of phrase repetition. To avoid this specific case, login and logout times were recorded for each subject, and a subject's data was discarded if that subject took less than the calculated amount of time required to complete each session.

RESULTS

As a preliminary test, the effects of session-to-session preference differences were analyzed. Using a repeated-measures ANOVA, no significant effects of the listening session were found ($F(2, 430) = 0.840$, $p = 0.433$). Overall, therefore, listener preferences did not significantly increase or decrease from session to session.

Since the hypotheses of this study are based on the inverted-U function between preference and subjective complexity, the relationship between these two factors was investigated next. Figure 2 shows a plot of the mean preference ratings for each subjective complexity rating as submitted during Session 1 of the study. Note that these preference ratings are song-independent, i.e. the same chorus typically receives different subjective complexity ratings by different subjects; therefore each point on the X-axis represents a combination of different songs, and each song is represented by multiple points on the X-axis.

Figure 2: Relationship between complexity and preference for chorus-only stimuli



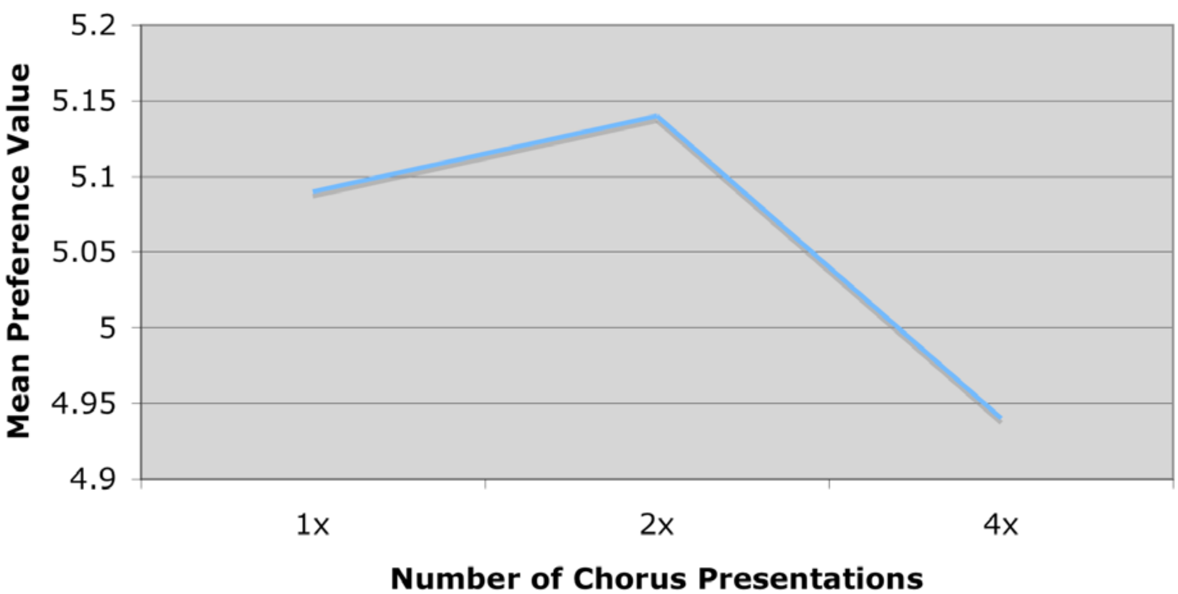
As Figure 2 shows, the overall relationship between subjective complexity and preference in this study was linear, not quadratic as the inverted-U theory predicts. This linear correlation found between preference and complexity was significant ($r = 0.390$, $p < 0.001$). It should be noted, however, that of the 270 chorus-complexity ratings (18 choruses * 15 subjects), only one had a value of 9. If this single value is ignored, a slight dip in the average preference rating from a complexity rating of 7 to 8 can be seen; however, this slight downward turn is not significant enough to imply the overall expected quadratic trend.

This disconnect between the inverted-U theory and the linear correlation shown above may initially seem to call into question the fundamental premise on which the hypotheses of this study are based. However, restitution can be found by positing that absolute subjective complexity ratings were never high enough to create decreases in preferences. In other words, none of the stimuli were so complex as to create higher-than-moderate levels of arousal. The linear correlation as shown in Figure 2, therefore, would roughly represent only the left-hand portion of the inverted-U curve, i.e. that portion in which preferences rise with subjective complexity.

To put it another way, the choruses overall were perceived as relatively simple. In fact, of the 270 complexity ratings, only 25% fell above the midpoint (within the range of 6, 7, 8, or 9). In contrast, 55% of the complexity ratings fell below the midpoint (i.e. had values of 1, 2, 3, or 4); the remaining 20% accounts for complexity ratings of 5. One subject (whose data was not used due to the subject's inability to the complete the study) mentioned, for instance, that she had specifically reserved (and thus not used) the higher end of the complexity scale since she expected more complex excerpts to follow.

Taking into account the limited slice of the inverted-U relationship that the stimuli explored, the expected effect of repetition on preference within the scope of this study becomes more straightforward. Specifically, a decrease in novelty (via repetition) should create an overall decrease in preference since the arousal levels of the stimuli (as shown via the complexity-preference ratings) appear to be at or below optimal levels. Figure 3 charts the overall median preference ratings for each level of repetition (across all stimuli). As this figure shows, the high-level repetition case (4x) does show a noticeable drop as compared to the other two cases (1x, 2x).

Figure 3: Overall relationship between preference and repetition level (all subjects)



Using a repeated-measures ANOVA, this overall effect of repetition level on preference was found to be significant ($F(2, 538) = 3.446$, $p = 0.033$). Although there is an obvious visual difference in Figure 3 between the 1x and 2x versions, it should be noted that this difference is only magnified by the Y-axis range. In particular, the difference in preference ratings between the 1x and 2x versions is not significant ($F(1, 269) = 0.388$, $p = 0.534$). The overall significant effect arises mainly from the differences between the 2x and 4x versions ($F(1, 269) = 6.980$, $p = 0.009$). These results – that a high level of repetition causes a decrease in preference – basically jibe with the experimental expectations.

The next step was to analyze potential interactions with repetition level as a determinant of preference. According to the inverted-U hypothesis, lower complexity ratings might produce more radical drops in preference when arousal is decreased since the quadratic function has a steeper slope at those points. Among all subjects, however, no significant effect of the interaction between repetition and complexity could be found ($F(16, 522) = 1.378$, $p = 0.147$). Table 2 below (in lieu of a colored-line graph) supports this lack of an interaction. For example, the highest-rated repetition level does not follow a clear trend. The pooling of complexity ratings in the 2-5 range does not help the resolution of this table.

Table 2: Mean preferences across repetition levels vs. initial chorus complexity ratings (includes all subjects)

		NUMBER of CH PRESENTATIONS				
		1x	**2x**	**4x**	**N**	**High**
COMPLEXITY RATING	**1**	3.89	4.44	3.56	9	2x
	2	3.57	3.57	3.64	42	4x
	3	5.40	5.07	4.95	43	1x
	4	4.98	5.05	4.95	55	2x
	5	5.07	5.26	4.83	54	2x
	6	6.38	6.00	5.72	29	1x
	7	5.85	6.42	6.08	26	2x
	8	5.91	6.27	6.36	11	4x
	9	8.00	8.00	8.00	1	n/a

Other possible interactions were also considered. While no significant effect of the interaction between repetition level and song could be found ($F(34, 504) = 0.769$, $p = 0.825$), there was a significant effect of the interaction between subject and repetition level ($F(28, 510) =$

2.065, $p = 0.001$). Upon closer inspection, it was found that 3 or 4 subjects displayed a strong conformance to the overall trend. These trend-setting subjects could all be categorized as members of the biased and/or trained participant groups. Therefore, the overall effect of repetition level was broken down into the effects on different subject groups. Biased cases showed a strong effect of repetition level on preference ($F(2, 106) = 9.207$, $p < 0.001$), whereas no such effect was found in the unbiased group ($F(2, 430) = 0.237$, $p = 0.789$). Similarly, the group of trained musicians displayed a strong effect of repetition level on preference ($F(2, 106) = 6.387$, $p = 0.002$), while no such effect was shown among untrained musicians ($F(2, 430) = 0.037$, $p = 0.964$). A summary of the overall average preference ratings for these groups is given in Table 3.

Table 3: Comparison of Mean Preference Ratings for Different Subject Groups

		NUMBER of CH PRESENTATIONS				
		1x	**2x**	**4x**	***F***	***p***
	All	5.09	5.14	4.94	1.378	0.033
SUBJECT GROUPS	**Biased**	5.28	5.44	4.31	9.207	< 0.001
	Unbiased	5.05	5.06	5.09	0.237	0.789
	Trained	5.04	5.30	4.33	6.387	0.002
	Untrained	5.11	5.10	5.09	0.037	0.964

Again, it should be noted that both the biased and trained groups only included three subjects each. Moreover, two subjects belonged to both groups, such that the two groups only differed by a single member. After dividing the participants into these different groupings, the interaction of complexity and repetition level was investigated once more. No significant effect of the interaction between complexity and repetition level with regard to preference ratings was found for any group, however: neither trained ($F(16, 90) = 0.851$, $p = 0.626$) or untrained musicians ($F(14, 416) = 1.079$, $p = 0.375$) nor biased ($F(16, 90) = 1.404$, $p = 0.158$) or unbiased subjects ($F(14, 416) = 0.691$, $p = 0.784$).

DISCUSSION

The results of this study imply that structural alterations to a piece of music can have a significant effect on listeners' preferences. Specifically, listeners can show sensitivity to the extent of phrase repetition within popular songs. This overall finding stands in contrast to the lack of alteration effects found in some earlier experiments. The effects of the current study, however, were limited to a small subset of the participants.

Among both the musically-trained and biased subject groups, a relatively dramatic drop in preference ratings – somewhere over a 10% decrease – occurred between stimuli with two presentations of the chorus versus four presentations. Since complexity ratings of the stimuli showed less-than-moderate arousal levels due to the observed linear correlation between complexity and preference ratings, this overall decrease is consistent with the predictions of Berlyne's two-factor arousal theory. Unfortunately, no consistent relationship existed between complexity ratings and changes in preference based on repetition level.

Within the musically-untrained and unbiased subject groups, however, no significant effect of repetition level on preference ratings could be found. As well, no interaction between complexity ratings and repetition level was evident with respect to preference ratings. The lack of effect among this larger subject group calls into question the applicability of the results. Since popular music is typically enjoyed by untrained and unbiased listeners, for example, should not the effects be clearly evident within this group if the results are to be meaningful?

A few reasons are plausible as to why effects were only noticed in the trained/biased groups. Foremost, trained/biased subjects may have shown significant preference changes since their perception was presumably heightened towards the structural effects under investigation. One trained subject (who was not a member of the biased group) mentioned that only after having completed the first couple listening sessions did he realize that the stimuli were being structurally altered. After this point of realization, the subject admits, he honed in on the effects of repetition and gauged his preference ratings on this criterion. It may be difficult, however, to untangle this potential "raised-awareness" from the more mundane effect of experimenter bias.

Untrained/unbiased listeners may thus not focus immediately on phrase repetition but rather on other factors. For instance, preference ratings may have been polarized by song-to-song interactions. In other words, a listener's preference for a particular artist or genre may be strongly fixed when contrasted against other artists or genres. Despite a potential change in preference based on repetition level, this change could be swamped by the surrouding material, especially when measured on a limited scale of 1-9.

Particular confounds in the stimuli themselves may have also overwhelmed any potential effects of phrase repetition. For example, when excerpts included fewer chorus iterations, more non-chorus material was presented. Was this non-chorus material more or less preferred than the chorus material itself? More importantly, repetition (whether harmonic, melodic, or textural) always occurs at some level between verse and chorus material as well as within the chorus itself. An examination of preference ratings based on phrase repetition ignores the extent to which repetition occurs on a more general level. How do we assign a quantitative value to repetition in separate domain in order to measure how much unity versus variety is happening at any given moment? This question gets at a problematic crux in theories of form and cannot be answered here.

Future Research:

In order to avoid potential song-to-song interactions, a more direct comparison method could be used. Since the strongest results in this study were noticed within the trained/biased groups, it appears that focusing the listener's attention on the effects under investigation is an important factor. A future experiment might thus present subjects with different versions of just the same song and then ask subjects to pick which version they prefer the most/least. This ranking system could be weighted by asking how much more or less the subject prefers a certain version to another. This "forced-choice" situation would have to account for order effects but would otherwise better control for preference changes based on unintended effects.

Another possible experimental revision might explore the entire range of possible subjective complexity. Since popular music excerpts appear to generally fall at or below moderate arousal levels, the full effect of the inverted-U relationship may be difficult to test without markedly different source material. Other genres of music, such as jazz, classical, or 20th-century art music, might provide a broader range of stimuli with which to study Berlyne's hypothesis. Because most listeners – even those dedicated popular music fans – are exposed to a variety of musical styles in everyday life (through radio, television, film, etc.), experimental work on Berlyne's theory may need to reflect this wide scope more faithfully.

APPENDIX

Table 4: Musical Examples Used in Stimuli Creation

Artist	Album	Song	Chorus Times Beg.	End	Tot.
Madonna	Like A Prayer	Cherish	0:37	0:44	0:07
Toto	Toto IV	Afraid of Love	1:18	1:26	0:08
Lucinda Williams	Car Wheels on a Gravel Road	Right in Time	0:43	0:51	0:08
Ron Sexsmith	Other Songs	Nothing Good	0:32	0:40	0:08
The Decemberists	The Crane Wife	The Crane Wife 3	1:37	1:45	0:08
Cat Power	What Would The Community Think?	Nude As The News	1:30	1:38	0:08
Halo Benders	God Don't Make No Junk	Don't Touch My Bikini	0:52	1:00	0:08
Liz Phair	Exile In Guyville	Never Said	1:01	1:10	0:09
Kanye West	Late Registration	Touch the Sky	0:56	1:05	0:09
Sly and The Family Stone	Anthology	Thank You	0:55	1:04	0:09
Avril Lavigne	Let Go	Mobile	0:43	0:53	0:10
Stevie Wonder	Innervisions	Too High	1:10	1:20	0:10
Def Leppard	Pyromania	Too Late For Love	1:25	1:35	0:10
Jay-Z	Vol. 3	S. Carter	1:11	1:22	0:11
The Velvet Underground	The Velvet Underground	Pale Blue Eyes	1:33	1:45	0:12
Dixie Chicks	Home	I Believe In Love	1:45	1:57	0:13
Creed	Human Clay	What If	1:10	1:23	0:13
Beck	One Foot In the Grave	Girl Dreams	0:53	1:06	0:13

All stimuli can be found at: http://www.midside.com/cog/examples/

BIBLIOGRAPHY

Batt, R. (1987). Comments on "The effects of instrumentation, playing style, and structure in the Goldberg Variations by Johann Sebastian Bach." *Music Perception*, 5 (2): 207-213.

Berlyne, D. E. (1971). *Aesthetics and psychobiology*. New York: Appleton-Century-Crofts. Quoted in Tan et al., 2006.

Bonds, M. E. (1991). *Wordless Rhetoric: Musical Form and the Metaphor of Oration*. Cambridge, MA: Harvard University Press.

Burnham, S. (2002). "Form." *The Cambridge History of Western Music Theory*, edited by Thomas Christensen. Cambridge: Cambridge U. Press.

Covach, J. (2005). "Form in Rock Music: A Primer." In *Engaging Music: Essays in Music Analysis*, edited by D. Stein. Oxford: Oxford University Press.

Deliège, I., Mélen, M., Stammers, D., and Cross, I. (1996). "Musical Schemata in Real-Time Listening to a Piece of Music." *Music Perception*, 14 (2): 117-160.

Gotlieb, H. and Konečni, Vladimir J. (1985). "The effects of instrumentation, playing style, and structure in the Goldberg Variations by Johann Sebastian Bach." *Music Perception*, 3: 87-102.

Hargreaves, D. J. (1984). "The Effects of Repetition on Liking for Music." *Journal of Research in Music Education*, 32 (1): 35-47.

Karno, M. and Konečni, V. J. (1987). "The Effects of Structural Interventions in the First Movement of Mozart's Symphony in G Minor K. 550 on Aesthetic Preference." *Music Perception*, 10 (1): 63-72.

Konečni, V. J. (1987). Response to Robert Batt. *Music Perception*, 5 (2): 215-217.

Laitz, S. G. (2003). *The Complete Musician: An Integrated Approach to Tonal Theory, Analysis, and Listening*. Oxford: Oxford University Press.

Meyer, L. B. (1989). *Style and Music*. Philadelphia: University of Philadelphia Press. Quoted in Tan et al., 2005.

Ockelford, A. (1991). "The Role of Repetition in Perceived Musical Structures." In *Representing Musical Structure*, edited by P. Howell, R. West, and I. Cross. London: Academic Press.

Orr, M. G. and Ohlsson, S. (2005). "Relationship Between Complexity and Liking as a Function of Expertise." *Music Perception*, 22 (4): 583-611.

Smith, J. D., and Melara, R. J. (1990). "Aesthetic preference and syntactic protypicality in music: 'Tis the gift to be simple." *Cognition*, 34: 279-98.

Stravinsky, I. (1942). *The Poetics of Music*. Translated by Arthur Knodel and Ingolf Dahl. Cambridge, MA: Harvard University Press.

Tan, S. and Spackman, M. P. (2005). "Listeners' judgments of the musical unity of structurally altered and intact musical compositions." *Psychology of Music*, 33 (2): 133-153.

Tan, S., Spackman, M. P., and Peaslee, C. L. (2006). Effects of repeated exposure on liking and judgments of thematic unity of patchwork and intact compositions. *Music Perception*. 23 (5): 407-421.

Tillman, B. and Bigand, E. (1996). "Does Formal Musical Structure Affect Perception of Musical Expressiveness?" *Psychology of Music*, 24 (3): 3-17.

Zajonc, R. B. (1968). "Attitudinal effects of mere exposure." *Journal of Personality and Social Psychology Monographs*, 9: 1-27. Quoted in Tan et al., 2006.