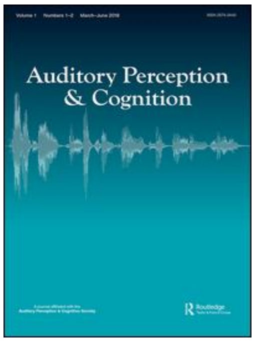

Auditory Perception & Cognition

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rpac20

Well-Formed Stimuli Lead to Perceptual Asymmetries in Discrimination: Evidence from Musical Chords and Rhythms

E. Glenn Schellenberg

To cite this article: E. Glenn Schellenberg (2021): Well-Formed Stimuli Lead to Perceptual Asymmetries in Discrimination: Evidence from Musical Chords and Rhythms, Auditory Perception & Cognition

To link to this article: https://doi.org/10.1080/25742442.2021.1886842

AUDITORY PERCEPTION & COGNITION
https://doi.org/10.1080/25742442.2021.1886842

Routledge
Taylor & Francis Group

ARTICLE

Well-Formed Stimuli Lead to Perceptual Asymmetries in Discrimination: Evidence from Musical Chords and Rhythms

E. Glenn Schellenberg[a,b]

[a]Instituto Universitário De Lisboa (ISCTE-IUL), Lisboa, Portugal; [b]Department of Psychology, University of Toronto Mississauga, Mississauga, Canada

ABSTRACT

In three experiments, listeners heard standard and comparison auditory sequences on each trial and judged whether they were the same or different. In Experiments 1 and 2, the sequences comprised chords (i.e., simultaneous combinations of pure tones) that were familiar (major), less familiar but with no sensory dissonance (diminished), or unfamiliar and dissonant. Performance was better in the major condition than in the other two conditions, but only when the major chord was the standard sequence. When it was the comparison, performance was poor. In Experiment 3, the stimuli were metrical or nonmetrical rhythms comprised of snare-drum beats. A discrimination advantage for metrical sequences was evident when the metrical sequence was the standard pattern but not when it was the comparison. In short, order of presentation determined whether well-formed stimuli facilitated discrimination. Well-formed auditory sequences led to advantages in discrimination when they were the standard (presented first), but this advantage was eliminated when the well-formed sequence was the comparison (presented second).

ARTICLE HISTORY
Received 21 September 2020
Accepted 31 January 2021

KEYWORDS
Music; discrimination; rhythm; perception; asymmetries

What makes a stimulus *well formed*? Well formed, here, refers to a stimulus that is easy to process and remember compared to other stimuli from the same class. Familiarity often plays a role (McFadden & Callaway, 1999), but some stimulus characteristics that afford goodness are almost certainly to be based on perceptual predispositions, such as those favoring symmetry in the visual domain (Palmer, 1991), and small-integer frequency ratios in the auditory domain (Schellenberg & Trehub, 1996b; Trainor, 1997). One consequence of well formedness is that a stable mental representation can be formed readily, such that alterations to the stimulus are easy to notice. In the present series of experiments, the question of interest was whether such enhanced discrimination of well-formed auditory stimuli would be affected by order of presentation.

Previous research in this area has focused on within-category discrimination of stimuli (e.g., numbers, countries, specific vowels), because between-category discrimination would be at ceiling levels provided the categories are psychologically relevant for the participant. In the auditory domain, well-formed stimuli have been described as perceptual *anchors*

CONTACT E. Glenn Schellenberg g.schellenberg@utoronto.ca Centro De Investigação E Intervenção Social (CIS-IUL), Instituto Universitário De Lisboa (ISCTE-IUL), Avenida Das Forças Armadas, Gabinete 2W9, Lisboa 1649-026, Portugal

(Acker, Pastore, & Hall, 1995), which lead to good discrimination, and as perceptual *magnets* (Kuhl, 1991), which lead to poor discrimination.

Acker et al. (1995) tested listeners' ability to detect slight alterations to prototypical and nonprototypical versions of a familiar musical structure – the major chord. *Chords* are simultaneous combinations of three or more tones, for which well formedness is determined in part by the intervals between adjacent tones, and whether such intervals are harmonic and/or familiar (Wild, 2002). Intervals in the major chord are both harmonic (or consonant) and familiar to individuals who have been exposed to Western music. On each trial in Acker et al.'s experiment, their listeners heard two pairs of chords, with both pairs comprising a standard chord followed by a comparison chord. The standard chord was the same for both pairs, but the comparison chord differed slightly for one pair, by mistuning the middle (*mi*) and/or the high (*sol*) tone by a small amount (≤0.33 semitone). The low tone (*do*) was fixed at 262 Hz (middle C). Listeners' task was to detect which pair (first or second) had a chord change. An anchor effect was confirmed by better performance when the standard chord was a prototypically tuned (equal tempered) major chord rather than a mistuned, nonprototypical variant of the same chord, but Acker et al. did not test for effects of presentation order.

In another study (McFadden & Callaway, 1999), the stimuli were six-tone chords that varied in well formedness. The well-formed stimulus was again the major chord. The poorly formed chord had much sensory dissonance, and comprised notes that did not come from the Western chromatic scale, thereby eliminating any effects of familiarity due to exposure to Western music. An AX (same-different) task required listeners to detect an alteration to a single tone in the stimulus chords. Discrimination was better for the well-formed chords, as it was for other familiar stimuli in additional experiments that tested detection of a change in duration to one tone in a familiar or unfamiliar melody, or the deletion of a band of frequencies in speech presented forward or backward. McFadden and Calloway used a fixed order of presentation, however, such that it remains unclear whether their well-formed stimuli would be as easily discriminated if they were the comparison pattern (X) rather than the standard (A) in an AX task. The role of dissonance is also unknown. In the experiment with chords, adjacent tones in the dissonant chord fell within the same critical band, such that amplitude fluctuations masked the individual frequencies of the tones. In other words, the results may be explained by the limitations of the human auditory system (Moore, 2012) and a poorly chosen chord for the comparison condition.

A different view of well-formed auditory stimuli comes from the perceptual magnet effect, which refers to prototypical speech sounds (Kuhl, 1991). Speech sounds such as vowels vary within and between speakers, yet listeners hear them as instances of the same vowel. In Kuhl's original study, stimuli comprised prototypical and nonprototypical versions of the vowel /i/ generated with a speech synthesizer. A go/no-go method tested listeners' ability to detect when a repeating *referent* vowel was substituted with a novel *comparison* variant of the same vowel. Adults and rhesus monkeys were required to press a button or lift a key, respectively, when they heard a vowel change. A conditioned head-turn procedure was used with 6-month-old infants. For all participants, positive feedback for correct responses was provided by way of a visual reinforcer (humans) or food (monkeys). Adult and infant listeners, but not monkeys, were better able to detect subtle changes to the vowel (by manipulation of the first and second formants) when the referent was the nonprototype rather than the prototype. Because similarity space around the prototype was contracted, it was said to act like a magnet. In a follow-up study (Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992), American infants showed a perceptual magnet effect for /i/, a vowel used in English but not in Swedish, whereas Swedish infants showed a magnet effect for /y/, a vowel in Swedish but not in English. Neither group of infants showed a magnet effect for the nonnative vowel. Presumably, exposure to one's native-language environment shrank dissimilarity space for the infant, such that different tokens of a native vowel were perceived as instances of the same vowel, which would aid language development. This perspective implies that the magnet effect should be independent of presentation order, although this hypothesis was not tested.

Discrimination asymmetries have been observed, however, in other experiments that involved musical stimuli. For example, when a musical key is established and listeners are asked to discriminate standard and comparison sequences of chords, performance is better if the comparison sequence contains a chord that is relatively unstable in the context, compared to instances when the unstable chord is in the standard sequence (i.e., when the comparison and standard chord sequences are reversed, Bharucha & Krumhansl, 1983). Similarly, when listeners are asked to discriminate *melodies* (i.e., sequences of tones rather than chords), performance is better when the standard melody is scalar (i.e., all tones from the same scale) and the comparison melody contains a nonscalar or relatively unstable tone in the established key, compared to when the standard and comparison are simply reversed (Bartlett & Dowling, 1988; Bharucha, 1984). Such effects are not due to the memorability of the scalar melodies, demand characteristics, or the number of 1-semitone intervals in the melodies (Bartlett & Dowling, 1988). Because scalar melodies conform to a Western scale, they would sound well formed (or more musical) to the participants, whereas the nonscalar melodies would sound more like random or incomplete tone sequences.

Even two tones can be asymmetric in similarity space when they are presented in a musical context. For example, F# and G – two tones separated by one semitone – have different functions in the key of C major. After C itself, G is the most stable tone in a C-major context. G is also in the C-major scale and the C-major triad, whereas F# is nondiatonic, a note from outside the C-major scale. Krumhansl's (1979) listeners heard standard and comparison tones, separated by an intervening sequence of tones that conformed to the C-major scale, and judged whether they were the same or different. Performance was much better when the standard was diatonic (e.g., G) and the comparison was nondiatonic (e.g., F#) than when the standard was nondiatonic (e.g., F#) and the comparison was diatonic (e.g., G). Thus, because G is well formed or stable in a C-major context, it gave rise to asymmetric discrimination. In the present series of experiments, we sought to explore asymmetries in discrimination that depended solely on the degree to which a stimulus is inherently well formed.

When music-like stimuli comprise *intervals* (two tones) presented sequentially or simultaneously in the absence of an established key, infants, children, and adults detect alterations to the interval more readily when it is well formed (i.e., harmonic or consonant, e.g., perfect fourth, perfect fifth, octave) rather than poorly formed (e.g., tritone, major seventh, minor ninth) (Schellenberg, 2001; Schellenberg & Trehub, 1994, 1996a, 1996b; Trainor, 1997). When the well-formed interval is the comparison pattern, however, performance deteriorates markedly. In other words, well-formed musical intervals – presented in the absence of a musical context – lead to asymmetries in discrimination performance.

Similar effects have emerged with auditory patterns structured in time rather than pitch. For example, a rhythmic sequence can be described as well formed if it is metrical, such that the listener perceives an underlying beat or pulse. A disruption to a metrical rhythmic sequence leads to asymmetrical discrimination, with better performance when the metrical rhythm is presented as the standard rather than the comparison pattern in a same-different task (Bharucha & Pryor, 1986). The rhythmic stimuli used by Bharucha and Pryor were atypical, however, in the sense that they comprised tones with a clear pitch (square waves) rather than drumbeats, and standard and comparison sequences comprised only seven tones. In other studies of rhythm perception (Hébert & Cuddy, 2002; Hopyan, Schellenberg, & Dennis, 2009), discrimination was better when a metrical rather than a nonmetrical rhythm was the standard pattern in an AX discrimination task, but the researchers did not examine the possibility of asymmetric discrimination.

In the present study, we sought to determine whether discrimination asymmetries extend to musical stimuli more generally. Our stimuli were less impoverished than those used previously in studies that reported discrimination asymmetries, specifically chords comprised of six tones presented simultaneously in Experiments 1 and 2, and rhythms comprised of 10 snare-drum beats presented sequentially in Experiment 3. All stimuli were presented without effort to establish a musical key. The stimulus chords and rhythms came from previous research (Hébert & Cuddy, 2002; Hopyan et al., 2009; McFadden & Callaway, 1999), which documented superior discrimination performance for well-formed stimuli when they were the standard pattern in an AX task, presented before a potentially altered comparison pattern. Nevertheless, based on previous results with more impoverished stimuli (e.g., two pure tones, brief rhythms), discrimination performance was expected to suffer when standard and comparison patterns were simply reversed.

The first experiment was designed to replicate and extend the results of McFadden and Callaway (1999, Frequency Discrimination Experiment), using the same stimuli, an additional stimulus condition, and an additional manipulation that reversed the order of standard and comparison stimulus patterns. Experiment 2 sought to determine whether discrimination asymmetries would be evident when the task required the listener to rely solely on pitch relations. Finally, Experiment 3 was designed to test for discrimination asymmetries when stimuli were structured temporally, such that only well-formed rhythms had a clear meter or beat.

Experiment 1

The stimuli were identical to those used by McFadden and Callaway (1999, Frequency Discrimination Experiment) except (1) the to-be-discriminated change was fixed (i.e., 1 semitone) on all trials of all conditions, (2) standard and comparison target tones were reversed for half of the participants, and (3) a third condition was added. The third condition was based on the diminished chord, which is considered "unstable" (i.e., in need of resolution) according to music theory. It is also relatively unfamiliar in Western music, particularly in Western popular music. In the diminished chord, adjacent tones are separated by at least three semitones, which guaranteed the absence of sensory dissonance for pure tones in the pitch range of our stimuli (B_3, one semitone lower than middle C, was the lowest tone). By contrast, the dissonant chord was both dissonant *and* unfamiliar, in the sense that unlike the other two chords, listeners would not have heard it in Western music. In other words, the design teased apart dissonance and familiarity. The hypothesis was that discrimination performance would be best in the condition with the well-formed (major) chord, provided the well-formed chord was the standard, but not when it was the comparison.

Method

Participants

The listeners were 36 undergraduates registered in an introductory course in psychology. They were recruited without regard to music training and received partial course credit for their participation. On average, they had 3.7 years of music lessons (*SD* = 4.7, data missing for 5 participants).

Stimuli

Half of the participants (n = 18) were tested in the *usual* order, in which some of the stimuli were identical to those used by McFadden and Callaway (1999). A same-different (AX) task required participants to detect a pitch change to one tone in a chord comprised of six pure tones (e.g., D_4-$F\#_4$-A_4-D_5-$F\#_5$-A_5, see Figure 1). The target tone was the second to the highest, and identical across conditions: 740.0 Hz ($F\#_5$). On *different* trials, it was shifted down one semitone to 698.5 Hz (F_5). As shown in (Figure 1), in both standard and comparison patterns, the chord with the embedded target tone was preceded and followed by an identical chord, presented two semitones lower (e.g., C_4-E_4-G_4 -C_5-E_5-G_5). All chords were 300 ms in duration, with a 500-ms silent interval between patterns but no space between chords within a pattern. All chords were a major chord (as

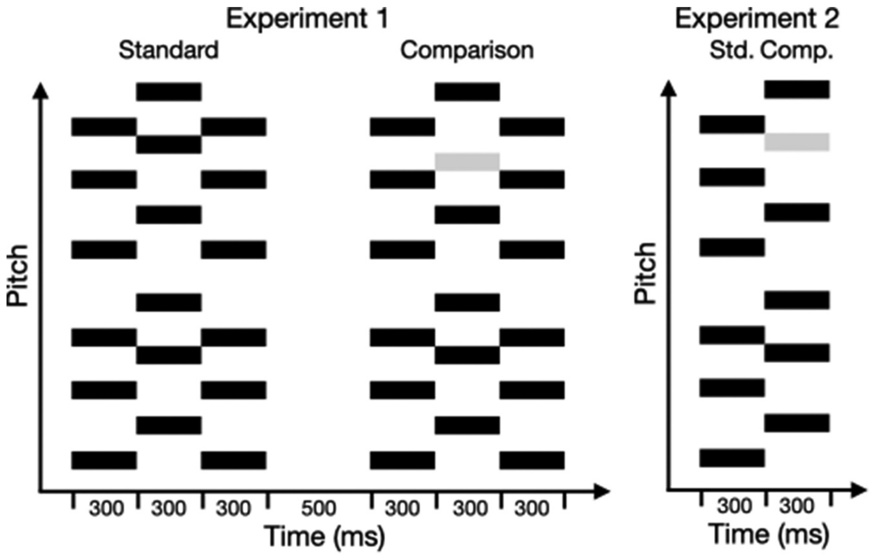

Figure 1. Schematic illustration of a trial in Experiment 1 and Experiment 2 (major condition, usual order). On *same* trials, standard and comparison patterns were identical in Experiment 1, and identical but transposed in Experiment 2. On a *different* trial, illustrated here, the gray tone was shifted down in pitch by 1 semitone relative to the standard.

in the above examples) in the *major* condition, a diminished chord in the diminished condition ($D\#_4$-$F\#_4$-A_4-$D\#_5$-$F\#_5$-A_5), or a highly dissonant chord in the dissonant condition. In the dissonant condition, the first (lowest), third, and fifth tones were drawn from the equal-tempered chromatic scale (i.e., D_4, $A\#_4$, $F\#_5$), but the second, fourth, and sixth tones were not. Rather, they were midway between $D\#_4$ and E_4, B_4 and C_5, and $G\#_5$ and A_5, respectively, such that intervals between adjacent tones could be as small as 1.5 semitones, and therefore highly dissonant.

To create additional variety in the stimuli, each condition had three versions, with chords presented in root position as well as in first and second inversions, as in McFadden and Callaway (1999), which created uncertainty regarding the key of the stimulus chords from one trial to the next. For example, in the major condition, the standard sequence had C major/D major/C major chords in root position, A major/B major/A major chords in first inversion, and E major/F# major/E major chords in second inversion. In each case, the potentially changed target tone was identical ($F\#_5$) and present in the middle chord, although its role in the chord varied from mediant (*mi* in D major) to dominant (*sol* in B major) to tonic (*do* in F# major), respectively.

The other half of participants ($n = 18$) were tested in the *reverse* order, in which the stimuli were identical but reversed, and the target tone was fixed at 698.5 Hz (F_5), except on *different* trials, when it was shifted upward by 1 semitone, to 740.0 Hz ($F\#_5$). Thus, in the major and diminished conditions, only the fifth chord in the sequence was actually a true major or diminished chord, respectively, and even then, only on *different* trials. All other chords (first, second, third, fourth, and sixth) had the second to the highest tone displaced downward by one semitone compared to the *usual* order. In the dissonant condition, chords were dissonant throughout, although the same manipulation was applied.

Procedure

Half of the listeners were tested in the usual order, the other half in the reverse order. Each listener was tested in three conditions (major, diminished, dissonant), with testing order counterbalanced, such that each of six possible orders had three participants. Listeners were tested individually in a sound-attenuating booth while sitting in front of a Macintosh computer wearing high-quality headphones. Before the test session began, an assistant used a version of the stimulus figure (Figure 1) as a visual aid to explain the task.

Trials were self-paced. On each trial, participants determined whether the target tone was identical in the second and fifth chords, by responding *same* or *different*. In each condition, the test phase was preceded by a 12-trial training phase to familiarize participants with the task and stimuli. In the training phase, the target tone was shifted upward by an octave (to 1480.0 Hz, $F\#_6$) on six *different* trials. The actual testing phase had 60 trials in each condition (30 *same*, 30 *different*). Stimuli were presented in five blocks of 12 trials each. In each block, each of the six possible stimuli (root position/first inversion/second inversion X same/different) was presented twice, with order randomized separately for each participant.

Results and Discussion

Hit and false-alarm rates were used to form three *d'* scores for each listener, one for each condition. Because performance was perfect for one listener in one condition (indeterminate *d'*), hit rates (responding "different" on *different* trials) and false-alarm rates ("different" on *same* trials) were transformed slightly by adding 0.5 to the numerator (the number of hits or false-alarms), and 1 to the denominator (the number of different or same trials) before calculating *d'*. This transformation – used for all participants throughout the three experiments – did not change the rank-order of *d'* scores (Thorpe, Trehub, Morrongiello, & Bull, 1988). The maximum possible *d'* score was 4.28.

Preliminary analysis confirmed that duration of music training was not significantly correlated with discrimination accuracy, in this experiment and in the ones that follow, so it was not considered further. Performance also did not vary as a function of testing order, so testing order was excluded from further consideration. As in the present experiment, testing order was counterbalanced with the stimulus conditions in Experiments 2 and 3.

The principal analysis was a mixed-design Analysis of Variance (ANOVA), with the condition-defining chord (major, diminished, dissonant) as a repeated measure, and stimulus order (usual, reverse) as a between-subjects variables. Descriptive statistics are illustrated in Figure 2. There was a significant two-way interaction, $F(2, 68) = 4.56$, $p = .014$, partial $\eta^2 = .118$, which motivated separate analyses of the three conditions using independent-samples *t*-tests. For the major condition, performance was markedly asymmetric, being better in the usual order than in the reverse order, $t(26.42) = 3.05$, $p = .005$, Cohen's $d = 1.02$ (unequal variances test), as predicted. For the diminished and dissonant conditions, order was irrelevant, $ps > .3$. The two-way interaction was also unpacked with separate examination of the two orders using one-way repeated-measures ANOVAs. For the usual order, performance varied across conditions, $F(2, 34) = 5.13$, $p = .011$, partial $\eta^2 = .232$, being better in the major condition than in the diminished, $p = .006$, and dissonant, $p = .031$, conditions, which did not

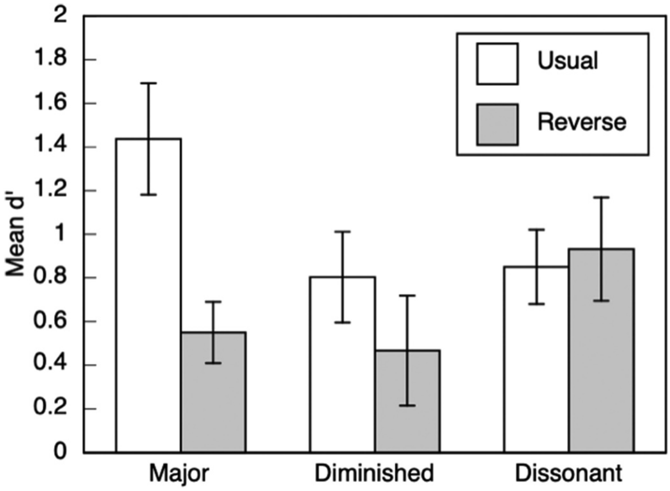

Figure 2. Descriptive statistics for Experiment 1, illustrated separately for condition (chord) and stimulus order. Error bars are standard errors. The effect of stimulus order was significant in the major-chord conditions but not in the other two conditions.

differ. For the reverse order, performance did not vary across conditions, $p > .1$. As shown in Figure 2, performance was best in the major-usual condition. In the major-reverse condition, however, performance was no better than it was in the condition with the worst performance (diminished-reverse), $p > .7$.

Thus, as in McFadden and Callaway (1999), who used the *usual* order, performance was better in the major-usual condition than in the dissonant-usual condition. Poor performance in the dissonant condition was not due solely to interference from sensory dissonance, because performance in the diminished-usual condition was similarly poor but worse than in the major-usual condition. In any event, the major chord conferred discrimination advantages when it was presented first in a same-different task, but not when it was presented second. Thus, discrimination involving a well-formed auditory stimulus (i.e., the major chord) was asymmetric.

Experiment 2

The goal of Experiment 2 was to test the generality of the findings from Experiment 1 using a procedure that was simplified, but a task that was more difficult. In the previous experiment, the first and third chords of the standard and comparison patterns may have contributed to response patterns. If so, the results could be specific to these six-chord contexts rather than a general perceptual phenomenon. Moreover, the task in Experiment 1 could have been solved, at least in principle, by detecting a frequency change to a single tone (740.0 Hz vs 698.5 Hz) that was presented in different contexts (i.e., chords), rather than (or in addition to) a change in the relations that define the goodness of the chord.

In the present experiment, listeners were again tested with a same-different (AX) discrimination task. On each trial, they heard two chords (standard and comparison) and judged whether the chords were the same or different. The standard and comparison chords were always presented in transposition, presumably making the task more difficult than in Experiment 1. Specifically, on all trials (*same* and *different*), all component tones were shifted in pitch from the standard to the comparison chord, such that the task could not be completed successfully by simply noticing a change in pitch. Rather, participants were required to determine whether the *relations* between component tones were the same or different, as in previous studies with musical intervals (Schellenberg, 2001; Schellenberg & Trehub, 1994, 1996a; Trainor, 1997). For the well-formed major chord, discrimination was expected to vary as a function of whether it was the standard or comparison chord. For the diminished and dissonant chords, no asymmetries were expected.

Method

Participants

The listeners were a new group of 36 undergraduates recruited as in Experiment 1. On average, they had 1.7 years of music lessons (*SD* = 3.2, data missing for 1 participant).

Stimuli

The stimuli comprised the fourth and fifth chords of the test sequences from Experiment 1 (i.e., the first and second chords of the comparison patterns, see Figure 1). On each trial, listeners heard a single standard chord followed by a single comparison chord. For half of these sequences (*same* trials), the second (comparison) chord was the same as the first (standard), except transposed upward by two semitones. For the other half (*different* trials), five of six chord tones were treated identically, but the remaining tone (the second to the highest tone of the comparison chord) was mistuned downward in the usual order, from 740.0 Hz (F#$_5$) to 698.5 Hz (F$_5$), or upward in the reverse order, from 698.5 Hz (F$_5$) to 740.0 Hz (F#$_5$), as in Experiment 1.

Procedure

The participants were divided evenly between the usual and reverse orders (*n*s = 18). The procedure was identical to Experiment 1, except that listeners heard only two chords on each trial, presented in transposition. Their task was to judge whether the two chords were the same or different. The initial practice trials demonstrated to participants that on *same* trials, the comparison chord was an exact transposition of the standard chord, but not the exact same chord.

Results and Discussion

Performance as measured with *d*' was again analyzed with a mixed-design ANOVA that had one repeated measure (condition: major, diminished, dissonant) and one between-subjects variable (stimulus order: usual, reverse). Descriptive statistics are illustrated in Figure 3. The two-way interaction was significant, $F(2, 68) = 12.82$, $p < .001$, partial $\eta^2 = .274$. Separate examination of the three chord conditions revealed better performance in the usual than in the reverse order for the major condition, $t(28.03) = 3.56$, $p = .001$, Cohen's $d = 1.19$ (unequal variances test), but no order effect in the diminished, $p > .3$, or dissonant, $p > .1$, conditions. Separate examination of the two orders confirmed that for the usual order, performance varied across conditions, $F(2, 34) = 10.39$, $p < .001$, partial $\eta^2 = .379$, with better performance in the major condition than in the diminished,

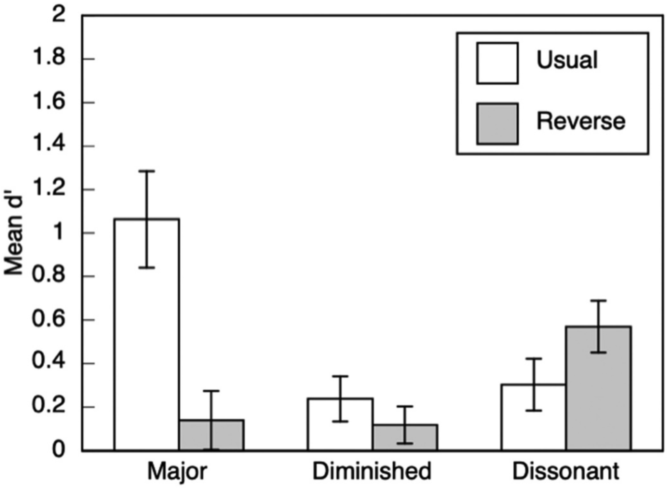

Figure 3. Descriptive statistics for Experiment 2, illustrated separately for condition (chord) and stimulus order. Error bars are standard errors. The effect of stimulus order was significant in the major-chord conditions but not in the other two conditions.

$p < .001$, and dissonant, $p = .008$, conditions, which did not differ, $p > .6$. For the reverse order, performance also varied across conditions, $F(2, 34) = 7.65$, $p = .002$, partial $\eta^2 = .310$, with better performance in the dissonant condition than in major, $p = .007$, and diminished, $p = .001$, conditions, which did not differ, $p > .8$. As in Experiment 1, although performance was best in the major-usual condition, in the major-reverse condition performance was no different than it was in the condition with the worst performance (diminished-reverse).

Thus, even when the task eliminated the possibility of using changes in absolute frequency as a cue, performance advantages in an AX discrimination task were evident when the well-formed major chord was the standard chord, presented first, but not when it was the comparison chord, presented second. When stimuli were less well formed (diminished and dissonant) chords, stimulus order did not matter.

Although Figures 2 and 3 indicate that performance deteriorated from Experiment 1 to 2, it is unclear whether response patterns stemming from the condition and order manipulations differed as well. The data sets were therefore combined and a new mixed-design ANOVA was conducted, which included Experiment as a third independent (between-subjects) variable. A main effect of Experiment confirmed that performance was poorer overall in Experiment 2 compared to Experiment 1, $F(1, 68) = 9.89$, $p = .002$, partial $\eta^2 = .127$. There was no two-way interaction between Experiment and condition, or between Experiment and order, and no three-way interaction, *F*s < 1, but the two-way interaction between condition and order remained strong, $F(2, 136) = 14.77$, $p < .001$, partial $\eta^2 = .178$. A marked asymmetry in performance due to order was evident in the major condition, $p < .001$, but not in the diminished or dissonant conditions, *p*s > .2. In short, although the task was more difficult in Experiment 2 compared to Experiment 1, the findings involving well formedness and presentation order were statistically identical.

Experiment 3

The goal of Experiment 3 was to determine whether discrimination asymmetries observed in Experiments 1 and 2 would extend to auditory sequences that had dynamic temporal properties but no clear pitches or pitch changes. On each trial, listeners heard two rhythms (i.e., sequences of drumbeats) and judged whether they were the same or different. As noted, for sequences of drumbeats, pattern goodness can be determined by the degree to which a meter (i.e., an underlying regular beat) is perceptible (Povel & Essens, 1985). Half of the stimulus rhythms were metrical, or *well formed*, because they had a clear beat (or meter). The other, non metrical rhythms had no clear beat. The hypothesis was that the metrical rhythms would confer good discrimination when they were the standard pattern in a same-different task, but not when they were the comparison pattern. For nonmetrical rhythms, we expected that asymmetrical discrimination performance would be eliminated or at least reduced.

Method

Participants

The participants were a new group of 36 undergraduates, recruited as in Experiments 1 and 2. They had 2.6 years of music lessons on average (*SD* = 3.6).

Stimuli and Apparatus

The testing apparatus was the same as in Experiments 1 and 2, but the stimuli were taken from Essens and Povel (1985), who created 35 rhythmic sequences that varied in the degree to which they implied a meter or beat. We used their five most and five least metrical rhythms, as in previous studies of meter perception (Hébert & Cuddy, 2002; Hopyan et al., 2009). In the five *metrical* rhythms, a beat in duple time was readily apparent to listeners. In the *nonmetrical* rhythms, a regular beat was difficult or impossible to perceive. Examples are illustrated in Figure 4.

Each of the 10 rhythms comprised 10 identical drum sounds presented with a snare-drum timbre (i.e., a digital sample from a Roland 808 drum machine). As with all percussive timbres, drumbeats had a rapid onset – reaching maximum amplitude almost instantly – followed by a decay of approximately 50 ms and a small amount of reverberation for an additional 100 ms. Each rhythm had the same duration: 3200 ms from the onset of the first drumbeat to the onset of the last drumbeat, which corresponded to two measures in 4/4 time (quarter-notes = 400 ms). Each rhythm also had the same nine intervals, or onset-to-onset durations between consecutive drumbeats: five of 200 ms, two of 400 ms, one of 600 ms, and one of 800 ms. The nine intervals were ordered differently across the 10 rhythms, except the 800-ms interval was always at the end (between the last two drumbeats). In the metrical but not in the nonmetrical rhythms, intervals with longer durations started on the downbeat (the first beat) of the two measures. Such intervals created a subjective accent (Povel & Okkerman, 1981), which gave rise to a perceived meter.

Procedure

Each participant was tested with metrical and nonmetrical rhythms. Both conditions were counterbalanced with testing order, and both had 60 trials: 30 same and 30 different.

Half of the 36 participants (*n* = 18) were assigned to the usual order. On each trial, they heard two rhythms (standard and comparison) separated by 2 s of silence. Each of the five metrical or nonmetrical rhythms was presented 12 times as the standard pattern (on six *same* trials and six *different* trials). On *same* trials, the standard and comparison were identical. On *different* trials, a silent gap of 100 ms was inserted to the end of the first of the two 400-ms intervals in the comparison pattern, thereby extending the duration of the pattern slightly (to 3300 ms). The other half of the participants (*n* = 18)

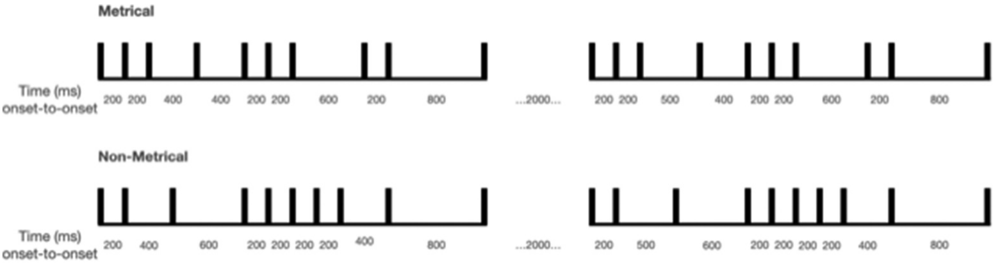

Figure 4. Schematic illustration of two different trials in Experiment 3. The upper and lower panel show metrical and nonmetrical rhythmic sequences, respectively. The comparison (right) differs from the standard (left) because it has an additional 100 ms of silence added to the first 400-ms interval.

were assigned to the reverse order, which was identical, except that the standard rhythm always included the additional 100 ms of silence. On *same* trials, the comparison rhythm was identical to the standard. On *different* trials, the additional 100 ms of silence was removed.

Results and Discussion

Two *d'* scores were calculated for each participant, one for metrical rhythms, another for nonmetrical rhythms. The principal analysis was a mixed-design ANOVA with one repeated measure: meter (metrical, nonmetrical), and one between-subjects variables: stimulus order (usual, reverse). There was a two-way interaction between meter and stimulus order, $F(1, 34) = 5.10$, $p = .030$, partial $\eta^2 = .131$. Descriptive statistics are illustrated in (Figure 5). In both conditions (metrical and nonmetrical), there was a discrimination advantage for the usual over the reverse order, but this simple effect was larger in the metrical condition, $t(20.30) = 3.31$, $p = .003$, Cohen's $d = 1.10$, than in the nonmetrical condition, $t(22.11) = 2.58$, $p = .017$, Cohen's $d = 0.86$ (unequal variances tests). For the usual order, performance was better in the metrical than in the nonmetrical condition, $F(1, 17) = 7.92$, $p = .012$, partial $\eta^2 = .318$. For the reverse order, performance was similar in both conditions, $p > .6$. Finally, performance was best in the metrical-usual condition, whereas performance in the metrical-reverse condition was no different that in the most difficult condition (nonmetrical-reverse), $p > .6$.

General Discussion

In three experiments, a same-different (AX) task was used to test participants' ability to discriminate well and poorly formed auditory stimuli from other stimuli that were altered slightly. Performance was much better when the well formed rather than the poorly formed

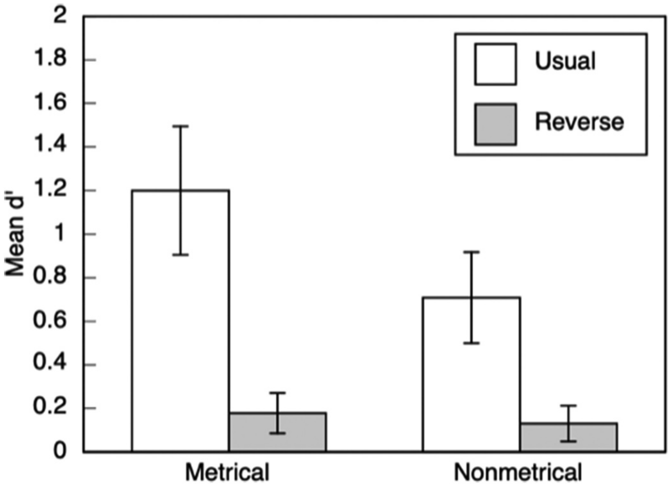

Figure 5. Descriptive statistics for Experiment 3, illustrated separately for meter and stimulus order. Error bars are standard errors. The magnitude of the stimulus-order effect was greater for the metrical conditions than it was for the nonmetrical conditions.

stimulus was the standard pattern. Performance deteriorated markedly, however, when the well-formed stimulus was the comparison rather than the standard, despite the fact that the two stimuli (A and X) were simply reversed in terms of order of presentation. Thus, when the test of discrimination included a well-formed stimulus, similarity space – as indexed by discrimination performance – was asymmetric. In fact, in all three experiments, any performance advantage for the well-formed stimulus was completely eliminated when it was the comparison pattern. Despite an increase in task difficulty from Experiment 1 to 2 due to the elimination of pitch-change cues, the magnitude of the asymmetry was identical for well-formed chords, and in Experiment 3, it extended to rhythms that were well formed in terms of temporal information.

Contrary to claims made by McFadden and Callaway (1999), then, well-formed (or *commonly encountered*) stimuli do not lead to "better discrimination" because this effect is limited to when they are the standard rather than the comparison pattern in same-different tasks. Accordingly, good discrimination is not necessarily the best marker of when a pattern has acquired well-formed (or commonly encountered) status. Rather, asymmetric discrimination appears to be key. Inspection of the large differences in height between the leftmost white and gray bars in the results figures (Figure 2, 3, and 5) highlight that this asymmetry effect was robust across analyses. Moreover, in each instance, the effect size (Cohen's *d*) was over 1 (i.e., performance differed between conditions by more than 1 *SD*). An unexpected result was that for each of these direct comparisons, Levene's test indicated that variances differed significantly between conditions, which motivated the use of separate-variances tests. As shown by the error bars in Figure 2, 3, and 5, variance was greater when the well-formed stimulus was the standard rather than the comparison. In other words, the reverse order with well-formed stimuli was difficult for most individuals, whereas better performance in the usual order was more variable across individuals. Future research could examine what individual-difference variables are predictive of this greater variability.

Across experiments, when the well-formed stimulus was the comparison pattern rather than the standard, discrimination performance was never any different, but never any worse, than it was in the condition with the worst performance. One would predict particularly poor performance, however, if the well-formed stimulus served as a perceptual magnet when it was the comparison pattern. As in Acker et al. (1995), then, well-formed auditory stimuli served as perceptually stable anchors rather than magnets, but only when they were presented first in a same-different task.

The reader may wonder why an order effect, albeit attenuated, was evident for the nonmetrical rhythms in Experiment 3. Even though there was no obvious quarter-note beat at 400-ms intervals, the nonmetrical sequences had a meter, at least in principle, at the 200-ms (eighth-note) level, although this may not have been perceived explicitly. In any event, the beat at the 200-ms level was disrupted in the altered sequences, such that the only possible meter was at the 100-ms (sixteenth-note) level, or 10 beats per s, which is too rapid to be a beat. (Imagine tapping your toe 10 times per s.) In other words, standard nonmetrical sequences were still better formed (i.e., more metrical) than altered comparison sequences, therefore conferring asymmetric discrimination performance, in line with the overall hypothesis of the present report.

The present results can also be used to generate hypotheses about studies that use operant-headturn or habituation methods with infants, or EEG with listeners of all ages. For example, as in Kuhl's studies (Kuhl, 1991; Kuhl et al., 1992), Schellenberg and Trehub (1996b, Experiment 1) used an operant-headturn procedure to test 9-month-old infants' ability to detect changes to harmonic pure-tone intervals that comprised either a well-formed (i.e., common across musical cultures), small-integer frequency ratio (3:2, 7 semitones, perfect fifth, or 4:3, 5 semitones perfect fourth), or a larger-integer ratio (45:32, 6 semitones, tritone). The lower tone was fixed at C_4. The higher tone of the referent interval was G_4, $F\#_4$, or F_4, in the 3:2, 45:32, or 4:3 condition, respectively, and mistuned downward by 250 cents for the comparison interval. Infants noticed this mistuning in the 3:2 and 4:3 conditions, but not in the 45:32 condition. The present findings suggest that if the referent and comparison intervals were switched, the advantage for small-integer frequency ratios would disappear. Another hypothesis stemming from the present results is that infant dishabituation would be more robust and more likely after habituation to a well-formed auditory or visual stimulus, compared to when the habituation and dishabituation stimuli are switched.

In studies that use EEG to test whether participants perceive changes to auditory patterns, a *mismatch negativity* (MMN) event-related potential (ERP) is often evident when an unexpected (oddball) pattern is substituted for the expected, repeating standard (Yu, Liu, & Gao, 2015). In other words, the MMN measures the perceived discrepancy between the sound and listeners' mental representation of what they expected to hear. In principle, this technique could be used to establish when a stimulus has well-formed status. For example, one would expect a larger MMN when the standard comprises properly tuned major chords and the comparison has a mistuned component tone, compared to when the comparison and standard chords are reversed.

The asymmetries in discrimination observed here parallel asymmetries in similarity that have been evident in nonmusical domains, with tasks that do not depend on perceptual acuity or fine-grained judgments. In Rosch's (1975) terminology, *cognitive reference points* refer to stimuli that are central (or prototypical) to a category, such as numbers that are multiples of 10, or horizontal (or vertical) lines. Such cognitive reference points lead to asymmetries in similarity judgments. For example, Rosch's participants were more likely to agree that 996 is essentially 1000 than 1000 is essentially 996, or that a line at 100° is basically 90° than a 90° line is basically 100°.

Perhaps asymmetric perceptual space for well-formed stimuli, such as those observed in the experiments reported here, is a general property of human perception and cognition that extends broadly across domains, modalities, and levels of analysis. Tversky (1977) proposed that *stimulus salience* accounts for the fact that "similarity is not necessarily a symmetric relation" (p. 333). According to him, prototypical stimuli are more salient than other stimuli, such that less salient stimuli are more similar to salient stimuli than vice versa. He documented that poorly formed geometric figures (line drawings) were more similar to well-formed figures than vice versa.

In the visual domain, Garner (1962) proposed that well-formed patterns are those that retain their identity after rotation and/or reflection. Examples include patterns that are shaped like an X or a cross. Garner found that well-formed visual patterns led to more efficient and rapid processing at encoding *and* retrieval (Garner & Sutliff, 1974; Sebrechts & Garner, 1981), and that they received higher subjective ratings of pattern goodness (Garner & Clement, 1963). These concepts, which rely on spatial symmetry, do not extend directly to most auditory patterns, which unfold over time. One exception is simple tone sequences that are redundant but symmetric about a point in time (e.g., C_4-E_4-G_4-E_4-C_4), which are processed more efficiently than asymmetric sequences with less redundancy (e.g., C_4-E_4-G_4-E_4-$C\#_4$). Older listeners have more exposure to music, however, such that familiarity plays a greater role and the effect becomes more culturally specific from infancy to childhood to adulthood (Schellenberg & Trehub, 1999).

As another example, consider amplitude (loudness, intensity). A louder tone is obviously more salient than a softer tone. When participants are asked to rate how much a tone changes in loudness (i.e., amplitude, intensity), increases in loudness receive higher ratings than decreases, even though the increases and decreases are identical in terms of magnitude (Neuhoff, 1998). Neuhoff speculated that the effect could be adaptive, because an approaching sound source is potentially a threat. The relative salience of increases over decreases in loudness is evident behaviorally and neuronally in mice and monkeys as well as humans (Deneux, Kempf, Daret, Ponsot, & Bathellier, 2016; Maier & Ghazanfar, 2007), and therefore consistent with the notion that events with increasing loudness are more likely to be a marker of approaching, potentially harmful interactions. In the experiments presented here, the well-formed patterns (major chords, metrical rhythms) may have been salient because they were easy to process and represent, unlike the poorly formed patterns. Superior short- and long-term memory for salient patterns may generalize broadly because salient patterns are less likely to represent random events but more likely to have ecological importance.

On an even higher-order cognitive level, the “return-trip effect” refers to the phenomenon that the time it takes to get someplace (the initial trip) typically seems longer than the time it takes to return home, even if the exact distance and time are the same. Familiarity (e.g., recognizing landmarks) does not appear to account for the effect, which remains evident if the return trip takes a different route with the same distance and time (van de Ven, van Rijswijk, & Roy, 2011). Van de Ven and his colleagues attribute the effect to expectancies. The initial trip has the goal of getting somewhere, which often takes longer than expected, primarily because people optimistically and systematically underestimate the duration of events, whether they are in the past or the future (Kahneman & Tversky, 1982; Roy & Christenfeld, 2007; Roy, Christenfeld, & McKenzie, 2005). The return trip, then, is expected to be longer in duration than it actually is, such that it seems relatively short. In Tversky’s terms, home would be more salient (and familiar) than the target destination.

In sum, the results presented here are consistent with those from other studies that used different methods and stimuli taken from multiple domains. In general, familiar, commonly encountered, symmetrical, referential, and/or well-formed stimuli give rise to asymmetries in similarity space. When Andy Warhol entitled his book, *The philosophy of Andy Warhol: From A to B and back again* (Warhol, 1975), one wonders whether Warhol was aware that getting from A to B was likely to take longer phenomenologically than getting back from B to A.

Disclosure Statement

No potential conflict of interest was reported by the author.

Funding

Funded by the Natural Sciences and Engineering Research Council of Canada, and the Fundação para a Ciência e a Tecnologia (FCT) em Portugal.

ORCID

E. Glenn Schellenberg http://orcid.org/0000-0003-3681-6020

References

Acker, B. E., Pastore, R. E., & Hall, M. D. (1995). Within-category discrimination of musical chords: Perceptual magnet or anchor? *Perception & Psychophysics*, *57*(6), 863–874.

Bartlett, J. C., & Dowling, W. J. (1988). Scale structure and similarity of melodies. *Music Perception*, *5*(3), 285–314.

Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition*, *13*(1), 63–102.

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology*, *16*(4), 485–518.

Bharucha, J. J., & Pryor, J. H. (1986). Disrupting the isochrony underlying rhythm: An asymmetry in discrimination. *Perception & Psychophysics*, *40*(3), 137–141.

Deneux, T., Kempf, A., Daret, A., Ponsot, E., & Bathellier, B. (2016). Temporal asymmetries in auditory coding and perception reflect multi-layered nonlinearities. *Nature Communications*, *7* (1), 12682.

Essens, P. J., & Povel, D.-J. (1985). Metrical and nonmetrical representations of temporal patterns. *Perception & Psychophysics*, *37*(1), 1–7.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: Wiley.

Garner, W. R., & Clement, D. E. (1963). Goodness of pattern and pattern uncertainty I. *Journal of Verbal Learning and Verbal Behavior*, *2*(5–6), 446–452.

Garner, W. R., & Sutliff, D. (1974). The effect of goodness on encoding time in visual pattern discrimination. *Perception & Psychophysics*, *16*(3), 426–430.

Hébert, S., & Cuddy, L. L. (2002). Detection of metric structure in auditory figural patterns. *Perception & Psychophysics*, *64*(6), 909–918.

Hopyan, T., Schellenberg, E. G., & Dennis, M. (2009). Perception of strong and weak meter rhythms in children with spina bifida meningomyelocele. *Journal of the International Neuropsychological Society*, *15*(4), 521–528.

Kahneman, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgments under uncertainty: Heuristics and biases* (pp. 414–421). Cambridge, UK: Cambridge University Press.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology*, *11*(3), 346–374.

Kuhl, P., Williams, K., Lacerda, F., Stevens, K., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science*, *255*(5044), 606–608.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, *50*(2), 93–107.

Maier, J. X., & Ghazanfar, A. A. (2007). Looming biases in monkey auditory cortex. *Journal of Neuroscience*, *27*(15), 4093–4100.

McFadden, D., & Callaway, N. L. (1999). Better discrimination of small changes in commonly encountered than in less commonly encountered auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(2), 543–560.

Moore, B. C. G. (2012). *An introduction to the psychology of hearing* (6th ed.). Bingley, UK: Emerald.

Neuhoff, J. G. (1998). Perceptual bias for rising tones. *Nature*, *395*(6698), 123–124.

Palmer, S. E. (1991). Goodness, Gestalt, groups, and Garner: Local symmetry subgroups as a theory of figural goodness. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The perception of structure: Essays in honor of Wendell R. Garner* (pp. 23–39). Washington, DC: American Psychological Association. doi:10.1037/10101-001

Povel, D.-J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*(4), 411–440.

Povel, D.-J., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics*, *30*(6), 565–572.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, *7*(4), 532–547.

Roy, M. M., & Christenfeld, N. J. S. (2007). Bias in memory predicts bias in estimation of future task duration. *Memory & Cognition*, *35*(3), 557–564.

Roy, M. M., Christenfeld, N. J. S., & McKenzie, C. R. M. (2005). Underestimating the duration of future events: Memory incorrectly used or memory bias? *Psychological Bulletin*, *131*(5), 738–756.

Schellenberg, E. G. (2001). Asymmetries in the discrimination of musical intervals: Going out-of-tune is more noticeable than going in-tune. *Music Perception*, *19*(2), 223–248

Schellenberg, E. G., & Trehub, S. E. (1994). Frequency ratios and the discrimination of pure tone sequences. *Perception & Psychophysics*, *56*(4), 472–478.

Schellenberg, E. G., & Trehub, S. E. (1996a). Children's discrimination of melodic intervals. *Developmental Psychology*, *32*(6), 1039–1050.

Schellenberg, E. G., & Trehub, S. E. (1996b). Natural musical intervals: Evidence from infant listeners. *Psychological Science*, *7*(5), 272–277.

Schellenberg, E. G., & Trehub, S. E. (1999). Culture-general and culture-specific factors in the discrimination of melodies. *Journal of Experimental Child Psychology*, *74*(2), 107–127.

Sebrechts, M. M., & Garner, W. R. (1981). Stimulus-specific processing consequences of pattern goodness. *Memory & Cognition*, *9*(1), 41–49.

Thorpe, L. A., Trehub, S. E., Morrongiello, B. A., & Bull, D. (1988). Perceptual grouping by infants and preschool children. *Developmental Psychology*, *24*(4), 484–491.

Trainor, L. J. (1997). Effect of frequency ratio on infants' and adults' discrimination of simultaneous intervals. *Journal of Experimental Psychology: Human Perception and Performance*, *23*(5), 1427–1438.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*(4), 327–352.

van de Ven, N., van Rijswijk, L., & Roy, M. M. (2011). The return trip effect: Why the return trip often seems to take less time. *Psychonomic Bulletin & Review*, *18*(5), 827–832.

Warhol, A. (1975). *The philosophy of Andy Warhol: From A to B and back again*. New York: Harvest/Harcourt.

Wild, J. (2002). The computation behind consonance and dissonance. *Interdisciplinary Science Reviews*, *27*(4), 299–302.

Yu, W., Liu, T., & Gao, D. (2015). The mismatch negativity: An indicator of perception of regularities in music. *Behavioural Neurology*, 469508. doi:10.1155/2015/469508.