Cognition 240 (2023) 105601

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Expectation adaptation for rare cadences in music: Item order matters in repetition priming

Aditya Chander [a,*], Richard N. Aslin [b,c]

[a] Department of Music, Yale University, 469 College St, New Haven, CT 06511, USA
[b] Child Study Center, Yale School of Medicine, 230 S Frontage Rd, New Haven, CT 06519, USA
[c] Department of Psychology, Yale University, 405 Temple St, New Haven, CT 06511, USA

ARTICLE INFO

Keywords:
Repetition priming
Music cognition
Expectation
Statistical learning
Bayesian inference

ABSTRACT

Humans make predictions about future events in many domains, including when they listen to music. Previous accounts of harmonic expectation in music have emphasised the role of implicit musical knowledge acquired in the long term through the mechanism of statistical learning. However, it is not known whether listeners can adapt their expectations for unusual harmonies in the short term through repetition priming, and whether the extent of any short-term adaptation depends on the unfolding statistical structure of the music. To explore these possibilities, we presented 150 participants with phrases from Bach chorales that ended with a cadence that was either a priori likely or unlikely based on the long-term statistical structure of the corpus of chorales. While holding the 50–50 incidence of likely vs. unlikely cadences constant, we manipulated the order in which these phrases were presented such that the local probability of hearing an unlikely cadence changed throughout the experiment. For each phrase, participants provided two judgements: (a) a prospective rating of how confident they were in their expectations for the cadence, and (b) a retrospective rating of how well the presented cadence matched their expectations. While confidence ratings increased over the course of the experiment, the rate of change decreased as the local probability of an unexpected cadence increased. Participants' expectations favoured likely cadences over unlikely cadences on average, but their expectation ratings for unlikely cadences increased at a faster rate over the course of the experiment than for likely cadences, particularly when the local probability of hearing an unlikely cadence was high. Thus, despite entrenched long-term statistics about cadences, listeners can indeed adapt to unusual musical harmonies and are sensitive to the local statistical structure of the musical environment. We suggest that this adaptation is an instance of Bayesian belief updating, a domain-general process that accounts for expectation adaptation in multiple domains.

1. Introduction

Repetition is a pervasive feature across various musical styles throughout the world, as well as across parameters such as harmony, melody, and rhythm. It offers listeners opportunities to encounter recurring patterns within a single piece and across multiple compositions (Margulis, 2014). Patterns that recur especially frequently across compositions are typically termed *schemata* (Gjerdingen, 2007) and listeners process them more easily than non-schematic structures due to their familiarity (Bharucha & Stoeckig, 1986; Tekman & Bharucha, 1998; Tillmann, Janata, Birk, & Bharucha, 2003). However, novel patterns that are repeated sufficiently within a piece can sometimes challenge the extensive schematic knowledge that listeners have acquired through enculturation to a particular musical style. This study aims to investigate whether Western listeners possess the capacity to adapt their expectations for non-normative harmonic patterns that deviate from a highly learned schema – namely, a *cadence* – if such patterns are extensively repeated in the short term during a single listening session. By exploring this phenomenon, we seek to gain insights into the dynamic processes underlying harmonic expectation adaptation and the interplay between short-term repetition and long-term enculturated musical knowledge.

One of the most important cognitive capacities that humans have is the ability to form expectations and make predictions about upcoming events, which we do in many sensory and cognitive domains (Bubic, Von Cramon, & Schubotz, 2010). There is evidence for expectation and

* Corresponding author.
E-mail address: aditya.chander@yale.edu (A. Chander).

https://doi.org/10.1016/j.cognition.2023.105601
Received 24 February 2023; Received in revised form 8 August 2023; Accepted 14 August 2023

prediction in the visual domain (Haith, Hazan, & Goodman, 1988; Kok, Jehee, & de Lange, 2012; Summerfield & Egner, 2009), in motor control (Körding & Wolpert, 2004), in decision making (J. I. Gold & Shadlen, 2007), and in auditory domains such as language (Altmann & Kamide, 1999; Federmeier, 2007; Kamide, 2008; Van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005) and music (Besson & Faïta, 1995; Huron, 2006; Jones, 2018; Meyer, 1957; Schmuckler, 1989; Steinbeis, Koelsch, & Sloboda, 2006). Expectation and prediction are valuable human capabilities in that they facilitate learning through response to errors (Frank, Woroch, & Curran, 2005; B. P. Gold, Pearce, Mas-Herrero, Dagher, & Zatorre, 2019; Keiflin & Janak, 2017). Additionally, multiple encounters with similar stimuli improve the encoding of such stimuli and allow for faster and more efficient processing (Kok et al., 2012; Logan, 1990; Wiggs & Martin, 1998). This process is known as *repetition priming*.

Accounts of repetition priming in the harmony perception literature fall into two categories. On the one hand, some scholars suggest that repetition priming (and priming more generally) is a process that implicates short-term sensory memory: specifically, if a target chord appears in a given harmonic context, that appearance has an effect on a listener's subsequent ability to process the target due to the acoustic image of the context overlapping with that of the target. However, the effects of this type of repetition on a listener's ability to process a chord are mixed, with repetition facilitating the perception of a target chord only if that chord did not provide harmonic closure to the phrase (Bigand, Tillmann, Poulin-Charronnat, & Manderlier, 2005). An earlier study showed that the inclusion of an unexpected target chord in the context did not facilitate its processing compared to a phrase that did not include it, with phrases ending with a typical closure chord receiving more accurate and faster responses (Bigand, Poulin, Tillmann, Madurell, & D'Adamo, 2003). The results of this study were not predicted by a computational model of auditory short-term memory (Bigand, Delbé, Poulin-Charronnat, Leman, & Tillmann, 2014) or a model combining representations of sensory similarity with a pretrained style-agnostic model of relationships between musical keys (Collins, Tillmann, Barrett, Delbé, & Janata, 2014).

On the other hand, Tillmann and Bigand (2010) articulate another possible definition of repetition priming that entails the repetition of an abstract structure over a longer timescale (on the order of minutes). This is consistent with more domain-general cognitive accounts of repetition priming, which suggest that it helps humans build implicit knowledge about the structure of their environment. Implicit knowledge is acquired through a process called *statistical learning*, in which distributional information is acquired through mere exposure (Aslin, 2017). A capacity present from infancy, statistical learning is used to segment continuous speech streams into words (Aslin, Saffran, & Newport, 1998; Graf Estes, Evans, Alibali, & Saffran, 2007; Isbilen, McCauley, & Christiansen, 2022; Saffran, Aslin, & Newport, 1996), segment tone streams into "tone-words" (Saffran, Johnson, Aslin, & Newport, 1999), detect non-adjacent statistical dependencies in streams of syllables (Newport & Aslin, 2004) and tones (Creel, Newport, & Aslin, 2004), acquire knowledge about an artificial musical grammar (Loui, Wessel, & Hudson Kam, 2010), and predict the order of presentation of visual scenes (Ellis et al., 2021; Fiser & Aslin, 2002). We adopt this more abstract account of repetition priming in our study because our aim is to determine how malleable long-term representations of musical cadences are when short-term deviations from canonical cadences are introduced.

While statistical learning is a powerful mechanism that generates expectations for the most likely outcome of an event, our expectations can be thwarted when the actual outcome is not the most probable. For instance, in language, the opening of a "garden path" sentence strongly implies a particular type of completion, but a less likely conclusion with a different grammatical structure follows. An example of such a sentence is "The horse raced past the barn fell", where the word "raced" would usually be interpreted as a main verb (as in "The horse raced past the barn") but is reinterpreted as belonging to a relative clause by the end of the sentence, with "fell" being the main verb. On average, readers take longer to process garden path sentences than regular sentences and make more mistakes in comprehending them (MacDonald, Just, & Carpenter, 1992; Waters & Caplan, 1996). The processing time deficit for garden-path sentences is exacerbated when presented simultaneously with unexpected musical syntax (Slevc, Rosenberg, & Patel, 2009). Another type of expectation mismatch in the linguistic domain is not grammatical but lexical, where the most expected word is replaced by a less likely word that has the same grammatical function and is still compatible with the prevailing discourse. An example of such a sentence is "At the grocery store he bought three eggs", where "eggs" is a less likely completion than, say, "apples" (as eggs typically come in cartons of six or more rather than being purchased individually) but is still consistent with the discourse of the sentence (as it is possible to buy eggs at the grocery store). Such replacements have been found to slow down reading times and elicit the N400 component in the EEG-evoked response when presented auditorily (Van Berkum et al., 2005). (The N400 component is typically associated with prediction errors in language; see Kutas & Federmeier, 2011 for a review.) These results suggest processing deficits for low-probability sentence completions.

Analogous situations may arise in the domain of music, which is characterised by probability relationships between sound objects (Meyer, 1957). Much work on musical expectation has focused on the perception of *cadences*. Cadences are formulaic patterns of *harmony* (chord motion) and *counterpoint* (the interaction of different lines in a musical texture) that end phrases in Western tonal music. They are used in a number of Western musical traditions, including both classical (Caplin, 2004; Harrison, 2020a; Neuwirth & Bergé, 2015) and popular styles (Sears & Forrest, 2021; Stephenson, 2002; Temperley, 2011). Cadences are thought to be the most predictable patterns in Western tonal music (Huron, 2006; Meyer, 1957), and this claim has been supported by corpus studies. In an analysis of 18th-century *chorales* (hymns in four parts) composed by Johann Sebastian Bach, Huron (2006) showed that the bigram probability of a cadence's final chord is on average more predictable than the bigram probability of a chord occurring mid-phrase. Sears, Pearce, Caplin and McAdams (2018) used the Information Dynamics of Music (hereafter IDyOM; Pearce, 2005, 2018) to model a corpus of 18th-century string quartets by Josef Haydn. IDyOM simulates statistical learning computations with variable-order Markov models across multiple *viewpoints* (features of the music such as pitch, melodic interval, duration, and metrical position). The resulting model estimates the *entropy* (uncertainty) and *information content* (surprisal) of musical events. Using IDyOM models defined by various combinations of three viewpoints (chromatic pitch, melodic interval, and scale degree), the authors found that terminal events of cadences were more predictable than those from non-cadential events, consistent with Huron (2006). Moreover, the models revealed prediction errors following the cadence, supporting the idea that a cadence is a moment of syntactic closure that does not strongly determine the subsequent musical material. Sears and Forrest (2021) also showed that IDyOM predicts similar syntactic closure effects at the final chord of cadential progressions in 20th-century popular music, since this chord is more predictable than either of the chords that surround it.

While cadences may be more predictable on average than mid-phrase musical events, there are many cadence types that occur with different probabilities. For instance, *perfect authentic cadences* (PACs) are the most commonly occurring cadence. They appear much more frequently than *deceptive cadences* (DCs), a family of cadences involving a deviation from the PAC schema in which the melody resolves to the same stable scale degree but the underlying chord is changed (Harrison, 2020a). In the Bach chorales corpus, phrases that ended with a PAC-compatible melodic pattern were almost 24 times as likely to end with a PAC than a DC (de Clercq, 2015). In the Haydn quartet corpus, PACs occur around 50% of the time and appear more than six times as often as DCs (Sears, Pearce, Caplin, & McAdams, 2018). Sears, Pearce, Caplin and McAdams (2018) used another IDyOM model with two viewpoints (scale degree and the set of intervals above the bass note) to predict the information content of terminal chords at DCs and PACs, finding that DCs had significantly higher information content than PACs. Indeed, they also found that IDyOM's predictions for the surprisal of the terminal event were consistent with music-theoretic models of cadential strength across multiple cadence types beyond PACs and DCs.

Sears, Pearce, Spitzer, Caplin, and McAdams (2019) showed that Western listeners are sensitive to these different occurrence probabilities, regardless of their level of musical training. In their study, musicians and nonmusicians rated phrases from Mozart piano sonatas ending with PACs as matching their expectations better than phrases ending with DCs. Their participants also reacted faster to intonation deviations at the final chord of these phrases (and more accurately in the case of nonmusicians) for PACs than DCs. The authors used IDyOM to model the behavioural responses to the cadence categories using variable-order models with three melodic viewpoints (chromatic pitch, melodic interval, and scale degree) and one harmonic viewpoint (the vertical interval set above the bass) to represent the musical surface. The information content estimates from these variable-order models correlated significantly with the expectation ratings and reaction times to intonation deviations for the different cadence categories, providing a better fit to the data than models representing the sensory similarity of harmonic contexts and targets. More generally, variable-order models have been found to correlate significantly and consistently with listeners' ratings of melodic phrase-ending unexpectedness and the entropy of these rating distributions, while models bounded at a low order (such as unigram or bigram models) did not exhibit consistent significant correlations (Hansen & Pearce, 2014).

Several other studies have manipulated PACs in more extreme ways to elicit surprise, replacing the final chord with another chord from a different key. These studies found different responses to unexpected chords, including slower judgements of the contour of a melody underpinned by these unexpected chords (Loui & Wessel, 2007), the elicitation of ERP components associated with surprise (Janata, 1995; Koelsch, Gunter, Friederici, & Schröger, 2000; Koelsch, Gunter, Wittfoth, & Sammler, 2005; Koelsch, Schmidt, & Kansok, 2002), and increased electrodermal activity (Steinbeis et al., 2006). Together, the results of these studies suggest a processing deficit for the low-probability cadence types. They support the idea that *schematic expectation*, where a mental template of the most likely outcome is compared to what actually happens in the music (Bharucha, 1987; Huron, 2006), dominates the perceptual experience of cadences.

The strength of schematic expectations demonstrated in several studies has led some researchers to conclude that priming a listener with repeated rare musical structures from a familiar musical style has no effect on their ability to process the event more efficiently. One important study that makes this claim is Tillmann and Bigand (2010), who passively exposed two groups of participants to chorale-like phrases ending in PACs before they completed a reaction-time task. In the exposure phase, one group only heard phrases that suddenly changed key for the cadence (the "less-related" condition, which occurs rarely in Western classical music), while the other group's phrases stayed in key for the cadence (the "related" condition, much more common in Western classical music). In the reaction-time task, participants responded as quickly and accurately as possible to the timbre of the final chord on trials containing either related or less-related phrase endings. The authors expected that the less-related group would exhibit an advantage for the less-related phrases in the test phase, but this was not the case: reaction times were still significantly faster on related endings in both groups. Only when the test and exposure items were identical did the difference between ending types become smaller in the less-related group, but these participants were still slower on the less-related endings.

The authors argue that schematic knowledge – which comes from years of listening experience dominated by the canonical PAC in the same key – overrode an adaptation to the true statistics of the music that the participants heard over the course of the experiment. To frame their argument using terminology first defined in Bharucha (1987) and adopted by others such as Huron (2006), Margulis (2005), and the authors themselves, *veridical* expectations – which are thought to be formed from specific experience with certain musical examples that may contradict schematic knowledge – did not facilitate the processing of those same events in the test phase. But even if their participants still responded slower to the less-related endings regardless of exposure type, they may still have adapted their expectations in the short term even if they did not respond faster overall to the low-probability targets. In other words, veridical knowledge may have still affected responses to the schematically rare events because of repetition priming occurring during the exposure phase, even if schematic expectations primarily explained the pattern of responses. In fact, Bharucha (1987) used DCs to illustrate the conflict between veridical and schematic expectations: a listener may know that a DC is coming based on prior experience hearing a piece, but the DC is veridically expected and schematically unexpected. However, since the analysis in Tillmann and Bigand (2010) averaged reaction times across trials, the timecourse of expectation adaptation was lost.

A different modelling approach that accounts for the timecourse of the experiment might have provided support for their claim. Such approaches have been used effectively to demonstrate syntactic adaptation in language. Fine, Jaeger, Farmer, and Qian (2013) found that reading times for garden-path sentences containing a low-probability relative clause decreased significantly faster over the course of the experiment than they did for garden-path sentences containing a high-probability main verb. They frame this adaptation in Bayesian terms: a corpus study estimated the prior probability of relative-clause resolutions (e.g. "The experienced soldiers warned about the dangers conducted the midnight raid") to be 0.8% and the probability of main verb resolutions (e.g. "The experienced soldiers warned about the dangers before the midnight raid") to be 70%. However, in their experiment, the two types of resolution were equally likely, so the surprisal for relative-clause resolutions sharply decreases and faster reaction times are predicted. Additionally, they found that presenting participants with an increased number of relative-clause resolutions before the presentation of the first main verb resolution resulted in slower reading times for the main verb, lending further support to the Bayesian adaptation framework where a higher surprisal is predicted at the first main verb. Myslín and Levy (2016) provide additional support for the Bayesian framework, finding that readers are also sensitive to higher-order statistics in the linguistic environment. They observed that garden-path sentences resolved with a sentential complement (e.g. "The reviewers acknowledged the study had been revolutionary") cluster together in natural language, and found that participants exposed to clusters of this syntactic structure processed them more rapidly than participants exposed to the same number spaced out in time. Finally, outside the domain of language, Desender, Donner, and Verguts (2021) asked participants to report how confident they were in which direction a collection of dots would move. These dots mostly moved randomly, but a variable proportion of them moved coherently to the left or right of the screen. They collected confidence ratings at the start of the trial and after a delay of one second where participants were presented with more dot motion (i.e., more evidence). After the presentation of additional evidence, confidence ratings increased if they were ultimately correct and decreased if they were ultimately incorrect. Together, these studies suggest that expectation is a time-dependent process and hence shows the importance of accounting for how the probabilistic structure of the environment changes over time.

In the musical domain, there is evidence of expectation adaptation for melodic structure. Musical expectation adaptation is typically modelled using IDyOM, which can be configured to encode both long-term knowledge (implicit musical knowledge acquired through extensive exposure to music over one's lifetime) and short-term knowledge (representing the acquisition of the local statistics of the musical environment). This combined model predicted listeners' expectedness ratings for artificially generated melodies better than a long-term model alone (Agres, Abdallah, & Pearce, 2018). In a more naturalistic setting, a combined long- and short-term IDyOM model also predicted the distribution of sung completions of incomplete melodies composed by a professional musician in a cloze task (Morgan, Fogel, Nair, & Patel, 2019). However, it is not known whether listeners can adapt their expectations for a priori rare *harmonic* patterns.

2. The present study

2.1. Aims and hypotheses

The purpose of the present study is twofold. First, we explore whether it is possible for listeners to adapt their expectations for rare cadences, so that by the end of the experiment they sound more expected than at the beginning. We measured expectation adaptation via two sets of explicit ratings: first, a participant's confidence in which cadence would end an incomplete musical phrase (the *completion confidence* or CC rating); and second, a retrospective rating of how well the cadence matched their expectations after hearing the complete phrase (the *expectation match* or EM rating). This design is similar to what was used in Experiment 1 of Sears et al. (2019), though we did not ask participants how specific their expectations were. We opted to use explicit ratings rather than an implicit measure such as reaction time or accuracy on a cover task. Implicit measures tend to be favoured in priming studies, since explicit ratings may not tap directly into implicitly acquired musical knowledge (Bigand, 2003), being unduly influenced by participants' reflection on the purpose of the task (Huron & Margulis, 2010; Isbilen, McCauley, Kidd, & Christiansen, 2017; Västfjäll, 2010). However, we found from pilot experiments using implicit measures that participants performed at ceiling and had high variability in their reaction times, resulting in poor sensitivity. This may have been resolved with in-person data collection using a low-latency interface for measuring reaction times, but the speed of data collection would have been considerably slower than what is possible with online recruitment platforms, particularly with the restrictions of the COVID-19 pandemic.

Participants provided ratings to phrases from Bach chorales that ended with either a conventional PAC or a DC featuring a particularly rare final chord. Details of this specific type of DC are discussed in section 3.2. We presented participants with equal numbers of PACs and DCs, a distribution that would not be predicted by the original Bach chorale corpus where PACs are far more common (Huron, 2006). We predicted that participants' CC ratings would decrease, reflecting the impact of hearing many DCs that conflict with their prior expectations for a PAC. We also predicted that participants would provide higher EM ratings for DCs by the end of the experiment than they did at the beginning, while PAC ratings would not significantly increase. This would indicate an adaptation to the local statistics of the musical environment. Both predictions are consistent with a Bayesian belief updating framework.

Second, we explore how manipulating the local probability of hearing a DC while maintaining the global statistical structure of the stimulus set (50% PACs, 50% DCs) affects the rate at which expectation ratings change. Given that the ratio of PAC to DC is high in the long-term statistics of Western music, a short-term exposure to 50% of each introduces a sudden and large discrepancy from the long-term distribution (i.e., the participant's prior expectation). Thus, we created three conditions in which the magnitude of this discrepancy from the prior expectation was varied, with DCs occurring on 20%, 50% or 80% of trials in a series of blocks presented one after another. The order of these blocks was either 20/80–80/20, 80/20–20/80, or 50/50–50/50 as detailed in section 3.4, thereby manipulating how rapidly the DCs were introduced while equating the overall DC statistics across blocks. Regardless of the order in which the blocks were presented, we computed the probability of hearing a DC at each trial using maximum likelihood estimation from all previous trials. We predicted that CC ratings would decrease at a faster rate if the local probability of hearing a DC is higher, as there is more conflicting evidence with participants' prior expectations. Additionally, we predicted that for a given trial, the rate at which EM ratings for PACs and DCs changed would respectively decrease and increase if the probability of hearing a DC at that trial also increased. Evidence in support of our hypotheses would provide a motivation for modelling ratings from trial to trial, an approach that has not been used in previous studies of expectation adaptation in the music domain and has thus led to the conclusion that long-term priors dominate expectations.

2.2. Modelling participants' responses

Although there are ongoing efforts to make publicly available versions of IDyOM to model harmony (P. Harrison, 2020), the current release of IDyOM (Version 1.6, 2020) is only configured to model *monophonic* music – that is, single musical lines with no accompaniment. It is possible to map hand-crafted features to arbitrary viewpoints in IDyOM, and this has been done in a few studies that measure or simulate expectancy (Cheung et al., 2020; P. Harrison, 2020; P. Harrison & Pearce, 2018; Sauvé & Pearce, 2019; Sears, Korzeniowski, & Widmer, 2018). However, there is no standard way to model harmony in IDyOM.

Since we presented our participants with Bach chorale phrases, we initially decided to use a variable-order IDyOM model trained on musical features that have been shown to be successful in estimating the key of a chorale (Quinn, 2010). These features are detailed in section 3.2. We opted to use these features in place of Roman numeral annotations to train our model, even though Roman numerals have been used to represent chords in several studies of harmonic priming and expectation (Bigand & Pineau, 1997; Koelsch et al., 2000; Poulin-Charronnat, Bigand, & Koelsch, 2006; Sears, Verbeten, & Percival, 2021; Sears, Verbeten, & Percival, 2023; Tillmann & Bigand, 2010). These studies make use of simplified chorale passages that stay in a single key with isochronous quarter-note chords and much of the voice-leading complexity removed. However, we used unmodified passages with a considerable degree of passing dissonance at the eighth-note level, with some phrases featuring modulations. We thus contend that it is not a trivial matter for a listener to assign Roman numerals to the chords in the chorale excerpts that we present, as there may be multiple possible interpretations that could result in different conclusions about the likelihood of certain chord transitions. Meanwhile, our representation does not rely on the specifics of the tonal interpretation, but leverages theories of voice leading at cadences (Caplin, 1998; D. Harrison, 2020).

Although using these features predicted higher surprisal at DCs than PACs, we were concerned that the difference in surprisal was overestimated due to the size of the chord vocabulary (a problem that would also arise if Roman numerals were used to represent the harmony). Therefore, while we still selected stimuli according to this IDyOM model, we instead opted to use *linear mixed models* to predict participants' responses to different cadences. Linear mixed models have been used successfully in linguistic expectation adaptation studies to model Bayesian belief updating (Fine et al., 2013; Myslín & Levy, 2016) as well as in studies of harmonic priming and expectation (Slevc et al., 2009; Wall, Lieck, Neuwirth, & Rohrmeier, 2020). We thus adopt a similar modelling approach here.

3. Materials and method

The following experiment was approved by the Institutional Review Board of Yale University (Protocol ID: 2000021951).

3.1. Participants

To ensure sufficient statistical power relative to previous studies examining the timecourse of expectation adaptation (e.g., Fine et al., 2013: 73 participants; Wall et al., 2020: 36 participants), we recruited

150 participants ($N = 150$; 73 female, 70 male, 7 did not specify) using the online platform Prolific (www.prolific.co). All participants were adults ($M = 35.59$, $SD = 12.98$) and lived in either the United States or United Kingdom. None reported having hearing difficulties. Participants were not screened on their self-reported musical ability or experience. Their scores on the combined Musical Training, Perceptual Abilities and Singing Abilities subscales of the Goldsmiths Musical Sophistication Index (GMSI; Müllensiefen, Gingras, Musil, & Stewart, 2014) ranged from 23 to 86 ($M = 48.7$, $SD = 12.3$) out of a maximum of 89. More details about participants' GMSI scores are reported in section B of the Supplementary Information. Eight additional participants completed the study but were rejected because they failed the attention check (see section 3.3 for details).

3.2. Stimuli

The stimuli were 50 chord sequences synthesised in a piano timbre. Each sequence was presented at a tempo of 100 quarter-note beats per minute (i.e., each quarter-note beat lasted 600 ms). The stimuli were initially rendered as MIDI files and converted to MP3 using the TiMidity++ (www.timidity.sourceforge.net) and FFmpeg (www.ffmpeg.org) command-line tools on a MacBook Pro running Mac OS X Version 12.4 (Monterey). The piano was rendered using TiMidity++'s default soundfont.

The chord progressions themselves were excerpts of 50 phrases taken from the complete Bach chorales, accessed via the music21 Python package (Cuthbert & Ariza, 2010). Each phrase was in 4/4 metre, lasted nine quarter-note beats in duration, and ended on a PAC. The final chord of the PAC in each phrase (indicated by a *fermata*) was always a major chord occurring on beat 1 of a measure. Its duration was adjusted to be one quarter note.

Similar to Ohriner (2013), the bass line's final note was required to be approached by an ascending perfect fourth or descending perfect fifth, which is standard for a PAC. We also imposed some constraints beyond those in Ohriner (2013) to ensure that major-key PACs were identified:

1. The penultimate chord was required to contain a simple or compound major third above the bass note. This interval could be realised by any one of the remaining three voices. This ensured that we did not include phrases with minor chords at the penultimate beat, as this would make a different type of cadence.
2. The soprano voice of the final chord was required to be the root of the chord; in other words, it had to have the same pitch class as the bass.
3. The phrase was required to be predominantly in the major mode. To determine this, the mode of each phrase was estimated with the Krumhansl-Schmuckler key-finding algorithm (Krumhansl, 1990). This algorithm finds the rotation of the major and minor probe-tone profiles from Krumhansl and Kessler (1982) that correlates most closely with the pitch content of the phrase. Any phrase estimated to be in one of the 12 minor keys was rejected.

The 50 selected excerpts were subsequently manually inspected to ensure they ended with a V(7)-I harmonic progression with both chords in root position, and with scale-degree 1 appearing in the highest sounding note in the final (target) chord. All excerpts passed this inspection. Subsequently, two phrases were set aside for the practice trials and another eight were reserved for foil trials. (The details of these practice and foil trials are described below in section 3.3.) Of the remaining 40 phrases, 20 of them were modified to end with a type of DC that is consistent with the definition of a DC in Caplin (1998) and D. Harrison, Bianco, Chait, and Pearce (2020). This DC ended on a diminished-seventh chord with the bass moving down by a semitone from the penultimate note and the soprano melody staying the same as it was for the PAC. An annotated example of an original PAC (showing where it meets the criteria for inclusion in the stimulus set) and its modified DC counterpart is shown in Fig. 1A, with the full PAC-ending phrase shown in Fig. 1B. A full corpus analysis of the Bach chorales revealed that the voice-leading pattern of this DC appears only twice in the 314 chorales in 4/4 metre, and these two instances are not cadential. Thus, we expected listeners' prior estimates of the probability of this cadence to be near zero, even though the continuation is musically lawful.

To select which 20 phrases to modify, we used IDyOM to train a model of the Bach chorales and used that model to estimate the difference in harmonic information content between the DC and the PAC for each of the 40 phrases reserved for the main task. This model was a variable-order multiple-viewpoint Markov model with an upper bound of five events and no incremental training on the test set (that is, we used IDyOM's long-term model, or LTM setting). The features in the model

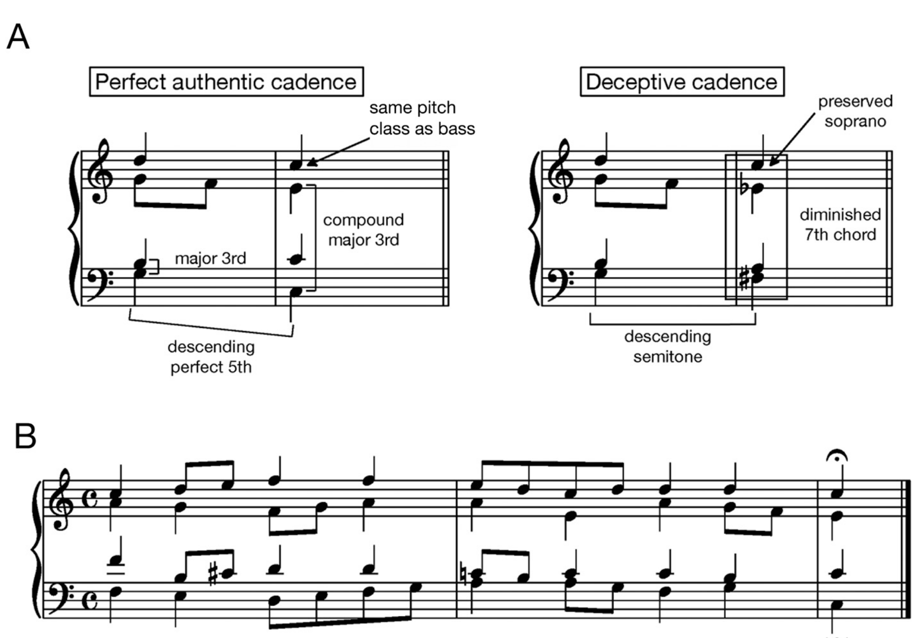

Fig. 1. A) A perfect authentic cadence (PAC) and a modified version showing the specific type of deceptive cadence (DC) used in this study. The PAC is defined by the features provided in the annotations: the bass descends by perfect fifth, the penultimate chord contains a major third above the bass, the final chord contains a (compound) major third above the bass, and the soprano and bass have the same pitch class (note name). The DC is also defined by the features provided in the annotations: the bass descending fifth is modified to be a descending semitone (half step), the final chord is changed to be a diminished seventh, and the soprano is the same as it is in the PAC. B) The complete phrase from the experiment that ends with the PAC illustrated in Fig. 1A.

were:

1. the interval between pitches in the bass line,
2. the set of harmonic intervals above the bass,
3. the ratio of chord duration between successive sonorities, and
4. the strength of the beat at which the chord appeared.

A combination of these features was used to predict the bass note and the chord quality. More details about the corpus modelling are available in section A of the Supplementary Information.

Subsequently, the 20 phrases with the highest estimated difference in surprisal between the DC and PAC were chosen for modification. For these 20 phrases, the average difference in surprisal between the DC and PAC was 16.72 bits (SD = 1.07 bits), whereas the average difference in surprisal for the remaining 20 phrases was 12.95 bits (SD = 2.40 bits).

3.3. Procedure

Participants accessed the experiment online using their personal computers. Those who signed up were redirected from Prolific to Gorilla (www.gorilla.sc), which was used to build and host the experiment (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020). They were instructed to use headphones and close all other tabs on their web browser. Participants were paid at the rate of $15/h through Prolific. After signing a digital informed consent form, participants' musical expertise was assessed using subscales of the Musical Training, Perceptual Abilities and Singing Abilities components of the GMSI. The specific questions that were asked are reported in section B of the Supplementary Information.

Participants were subsequently presented with instructions for the task, which was similar to Experiment 1 of Sears et al. (2019). On each trial, a stimulus was first presented without the final chord (i.e., just the first eight beats), and participants were instructed to rate how certain they were about what chord would come next on a scale of 1 (not confident at all) to 7 (extremely confident). This was the CC rating. Then, they heard the full sequence, and were asked to rate how well the completion matched what they expected to hear on a scale of 1 (not well at all) to 7 (extremely well). This was the EM rating. They were encouraged to use the full range of the 1–7 scale in their responses. During stimulus presentation, participants were instructed to focus their gaze on a loudspeaker icon that was presented in the middle of the screen. The sequence of screens displayed in a trial is shown in Fig. 2.

Participants received two practice trials before the main task in which both complete phrases ended with PACs. The main task consisted of 48 trials. Forty trials used the reserved PAC- and DC-ending phrases described above in section 3.2. The remaining eight trials were foil trials and consisted of the following:

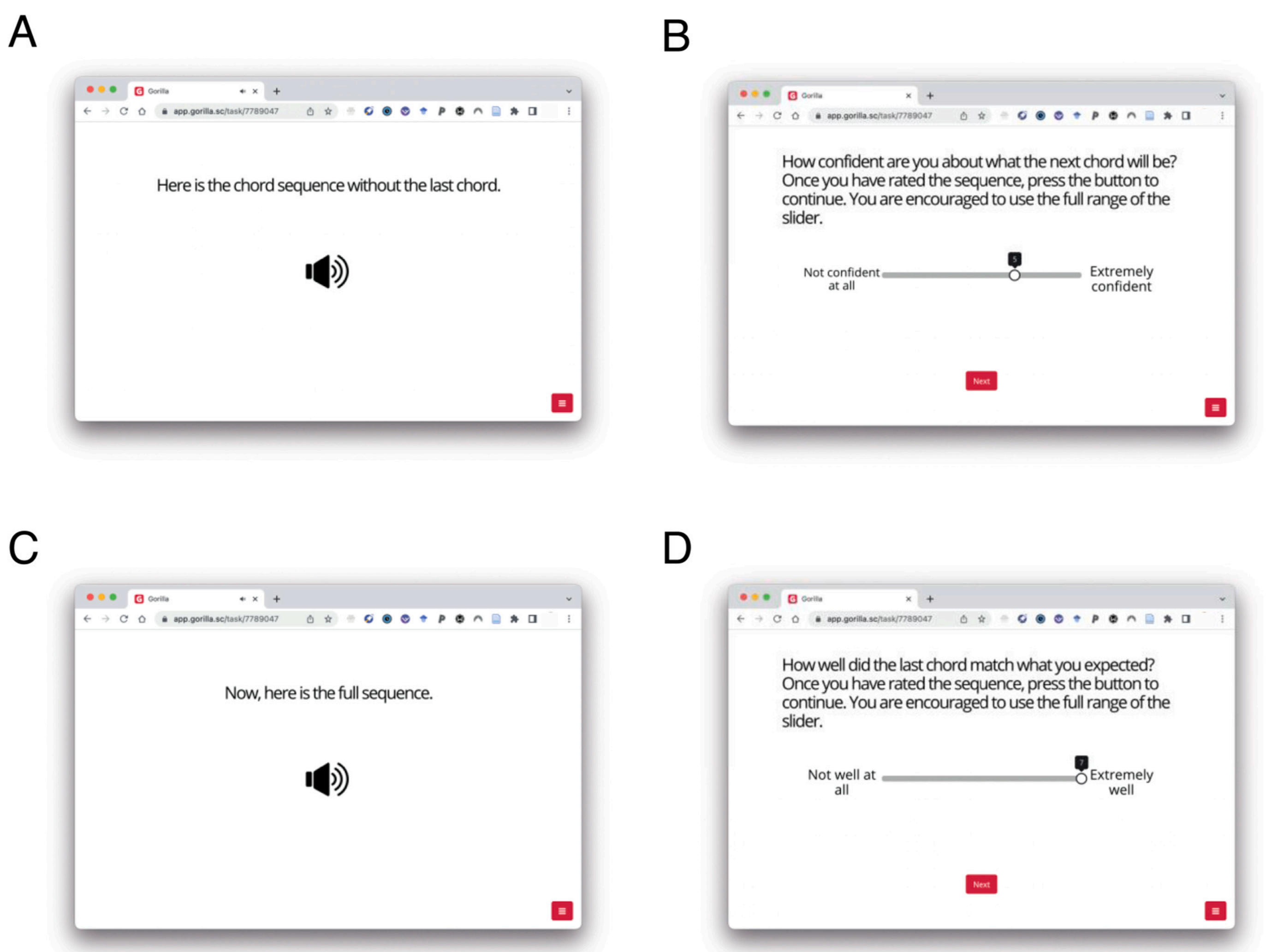

Fig. 2. The four screens presented to participants on the regular experimental trials. (A) Participants hear the phrase without the final cadential chord. (B) Participants rate how confident they are about what chord will follow. (C) Participants hear the full sequence. (D) Participants rate how well the final chord matched their expectations.

1. *Length foils.* On four trials, participants received the same pair of questions, but instead of hearing an eight-beat truncated stimulus followed by the nine-beat complete version, they heard just the first six beats as the truncated version followed by the first seven beats as the complete version. All length foils were also presented in a piano timbre. Length foils were included to make the task less monotonous, but also to provide a contrast with the cadential contexts by presenting phrase openings that were less predictive of what followed them.
2. *Timbre foils.* On another four trials, participants heard just the first six beats of a phrase. For two of these trials, the timbre of the last chord was changed to sound like an organ instead of a piano. Participants were instructed to report whether the instrument changed or stayed the same. They were not presented with a complete version of the stimulus after responding.

Performance on the timbre foil trials was used as a measure of attention, and participants were warned that they might occasionally receive a different question from what they were used to. If participants got fewer than three questions right, their data were excluded. Additionally, if a participant used fewer than three response values for the CC or EM ratings, their data were excluded for only that response category. Of the 150 participants who completed the experiment, one participant's CC ratings and another participant's EM ratings were excluded under this criterion.

3.4. Trial randomisation

The trials in the main task were split into two sub-blocks of 24 trials that followed each other without a break. Each sub-block contained two length foils, two timbre foils and 20 regular trials. There were three conditions, each corresponding to a different distribution of the PACs and DCs across the sub-blocks, with approximately equal numbers of participants assigned to each condition ($n_{50/50} = 51$, $n_{80/20} = 49$, $n_{20/80} = 50$):

1. *The "50/50" condition.* Both sub-blocks featured 10 DCs and 10 PACs.
2. *The "80/20" condition.* The first sub-block featured 16 DCs and 4 PACs, while the second featured 4 DCs and 16 PACs.
3. *The "20/80" condition.* The inverse of the 80/20 condition, the first sub-block featured 4 DCs and 16 PACs, while the second featured 16 DCs and 4 PACs.

Within each sub-block, the order of the 24 trials was randomised.

3.5. Analysis

Firstly, we checked whether participants were indeed providing lower average CC ratings for the length foils compared to the cadential trials. To test this, we used a paired-sample *t*-test on the mean length foil and mean cadential CC rating for each participant. For all other analyses, only data from cadential trials were analysed, as it was believed that phrases ending with a non-cadential harmony would not contribute to any change in expectation for a particular cadence.

3.5.1. Representing trials

Each cadential trial was represented using three features:

1. The *cadence*. This feature was sum-contrast coded with DC = 1 and PAC = −1.
2. The *proportion of prior non-foil trials ending with a DC* (hereafter *DC proportion*). This feature could take on any rational value from 0 to 1. In a Bayesian framework, it represents the posterior distribution of PACs and DCs at the current trial. Note that we did not model the condition directly as this would result in a loss of information about the true real-time statistics of the stimulus set.
3. The *item order*. This feature could take on any integer value from 1 to 40. It represents the index of the trial relative to all other non-foil trials. In a Bayesian framework, it is an estimate of the amount of evidence for the posterior distribution.

3.5.2. Modelling ratings for cadential trials

Considering just the cadential trials, there were 5960 observations for each rating type (40 ratings × 149 participants). We modelled CC and EM ratings separately using linear mixed-effects models fitted with the *lmer* function from the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015) in R. All models were fitted using maximum likelihood estimation to be able to make model comparisons.

For the CC ratings, we modelled responses in two ways. To observe the overall change in confidence regardless of the local statistics, we built a simple model with a fixed effect of *item order*, a random intercept for the different stimuli, and a by-participant random intercept and slope for *item order*. Then, to account for the local statistics, we built a second full model that added an interaction term between *item order* and *DC proportion*. (The effect of *DC proportion* on its own was excluded, as without *item order* it would represent an estimate of the prior probability of DCs, but this probability is already estimated to be zero.) We also added a by-participant random slope for the *item order* × *DC proportion* interaction, and as before we included random by-participant and by-stimulus intercepts.

For the EM ratings, we also modelled responses in two ways. Firstly, to observe the overall expectation adaptation regardless of the local statistics, we built a simple model with fixed effects of *item order*, *cadence* and their interaction, a random intercept for the different stimuli, and by-participant random intercept and slopes for all fixed effects and interactions. The second full model featured the same fixed effects and interactions, plus the two-way interaction of *item order* × *DC proportion* and the three-way interaction of *cadence* × *item order* × *DC proportion*. (The effect of *DC proportion* on its own was excluded, as was its interaction with *cadence*, for the same reason as the choice to exclude *DC proportion* from the second CC model.) The by-participant random slopes for *item order* × *DC proportion* and *cadence* × *item order* × *DC proportion* were also added, and as before we included random by-participant and by-stimulus intercepts.

The significance of the fixed effects was determined by analyses of variance (ANOVA) using the Satterthwaite method to approximate the degrees of freedom with the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2017). Additionally, for each rating type, we compared the fit of the simple and full models against each other and against a null model containing only by-participant and by-stimulus random intercepts. Following Nakagawa and Schielzeth (2013), we calculated the marginal R^2 (variance explained by fixed effects only) and conditional R^2 (variance explained by both fixed and random effects) for the three CC and the three EM models to provide an estimate of the total variance explained by each model. Then, we used chi-square tests to measure whether the full models provided a significantly better fit to the data than the simple models and null models. Finally, we compared the tradeoff between model fit and model complexity using the Bayesian Information Criterion (hereafter BIC; Schwarz, 1978). The formula for BIC is

$$\text{BIC} = -2 \times LL + d \times log(N)$$

where LL is the log likelihood of the model, d is the number of parameters in the model, and N is the number of observations. Models with a smaller BIC are typically preferred.

For both CC and EM ratings, we also performed moderation analyses to see whether the effects were driven by musical sophistication, indexed by the GMSI scores. These analyses are reported in section C of the Supplementary Information. The moderation analyses revealed that the key results were identified in subgroups with both above-average and below-average subgroups GMSI scores, so we will not discuss them further.

4. Results

4.1. Foil vs. non-foil trials

Average CC ratings for cadential trials ($M = 4.84$, $SD = 0.92$) were significantly higher than those for length foils ($M = 4.09$, $SD = 1.06$), $t(148) = 10.46$, $p < .001$. The effect size was large, with Cohen's $d = 0.86$.

4.2. CC ratings

4.2.1. Ignoring local statistics

The simple CC model's coefficient estimates and analysis of variance are presented in the upper half of Table 1. The main effect of *item order* was significant. The positive coefficient means that the confidence in how the phrase will end increases over the course of the experiment when not accounting for the local statistics of the stimulus set at the current trial.

4.2.2. Incorporating local statistics

Fig. 3 displays line plots of the full linear mixed model's estimates of CC ratings over the non-foil trials at different values of *DC proportion*. The model's coefficient estimates and analysis of variance are presented in the lower half of Table 1. Both the main effect of *item order* and the *item order* × *DC proportion* interaction were significant. This means that at a baseline level of *DC proportion* = 0, the confidence in how the phrase will end increases over the course of the experiment. However, as *DC proportion* increases, the rate of change of the CC ratings decreases. At a value of *DC proportion* = 0.74, the slope of the CC ratings dips below 0 and the confidence in how the phrase ends becomes progressively lower over the course of the experiment.

4.2.3. Model comparison

Table 2 contains the in-sample model fits and model comparisons for the null, simple, and full CC models. While the null CC model had a marginal R^2 (proportion of variance explained by fixed effects only) of 0, since it contains no fixed effects, the low marginal R^2 of the simple and full CC models suggests that the fixed effect of *item order* and the interaction *item order* × *DC proportion* do not explain much variance on their own, despite the high significance of their coefficients.

The null model's conditional R^2 (proportion of variance explained by fixed and random effects) was moderate, but the conditional R^2 values for the simple and full models were considerably higher than those for the null model, suggesting that the explanatory power of the model improved with the addition of by-participant random slopes. However, the conditional R^2 values of the simple and full models are very close to each other.

Table 1
Coefficient estimates and analysis of variance for fixed effects in linear mixed effects models predicting completion confidence ratings.

Predictor	Estimate	*SE*	*df*	Wald *F*	*p*
Simple model (ignoring local statistics)					
Item order	0.01	<0.01	165.47	17.96	<0.001***
Full model (incorporating local statistics)					
Item order	0.03	0.01	1624.23	40.50	<0.001***
Item order × DC proportion	−0.04	0.01	1210.47	22.20	<0.001***

Note. $N = 5960$ (40 non-foil trials × 149 participants). SE = standard error, df = denominator degrees of freedom for Type III Wald F tests (which were determined with the Satterthwaite approximation). *** $p < .001$.

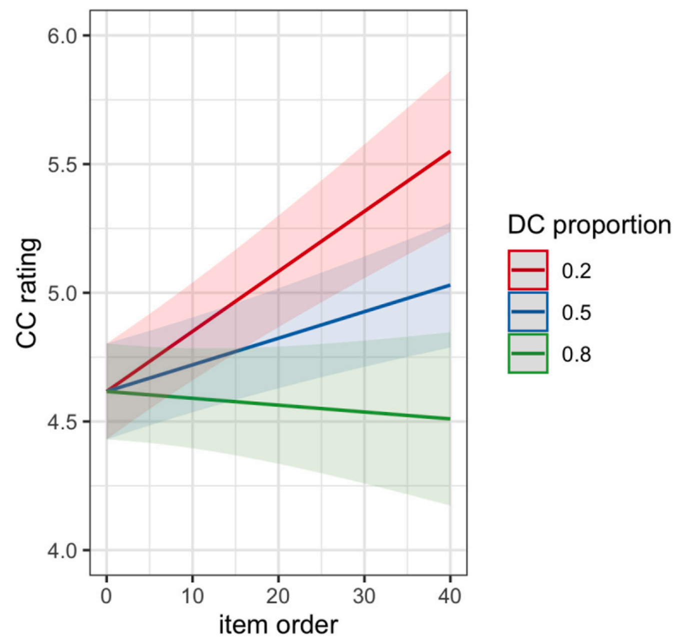

Fig. 3. Completion confidence (CC) ratings estimated by the full linear mixed model that accounts for the local statistics of the stimulus set. Slopes are shown for three different values of *DC proportion* (0.2, 0.5 and 0.8).

Unsurprisingly, the simple model explained significantly more variance than the null model. However, despite their seemingly similar model fits, the full model explained significantly more variance than the simple model according to the χ^2 test. Additionally, the full model had a lower BIC than the simple model (which in turn had a lower BIC than the null model), meaning that the inclusion of predictors that capture the local statistics of the stimulus set improved the tradeoff between model fit and model complexity.

4.3. EM ratings

4.3.1. Ignoring local statistics

The simple EM model's coefficient estimates and analysis of variance are presented in the upper half of Table 3. The main effects of *item order* and *cadence* were significant. The negative *cadence* coefficient means that DCs were rated as less expected on average than PACs, while the positive *item order* coefficient means that EM ratings increased over the course of the experiment. The interaction was not significant, meaning that there was no difference in the rate of change of EM ratings when not accounting for the local statistics of the stimulus set.

4.3.2. Incorporating local statistics

Fig. 4 displays line plots of the full linear mixed model's estimates of EM ratings over the non-foil trials at different values of *DC proportion* for the two different cadence types. The model's coefficient estimates and analysis of variance are presented in the lower half of Table 3. The main effect of *cadence* was significant, meaning that before a participant even completes any trials, they would theoretically rate a PAC as matching their expectations better than a DC. While the effect of *item order* was not significant on its own, the interaction between *item order* and *cadence* was significant. The interpretation of this coefficient is that the ratings of DCs decreases over the trials at a faster rate than PACs at a baseline level of *DC proportion* = 0. The interaction between *item order* and *DC proportion* was significant, meaning that ratings on average increased at a greater rate as *DC proportion* increased. However, the significant three-way interaction between *cadence*, *item order* and *DC proportion* shows that the increase was greater for the DCs than for the PACs. Indeed, looking at Fig. 4, the model estimates that EM ratings for DCs decrease

Table 2
In-sample model fits and model comparisons for null, simple and full models of completion confidence ratings.

Model	Marginal R^2	Conditional R^2	Number of parameters	BIC	χ^2 value (df)[a]	p
Null	0	0.37	4	19,796	–	–
Simple	0.01	0.49	6	19,599	215.25	<0.001***
Full	0.01	0.50	8	19,589	27.36	<0.001***

Note. $N = 5960$ (40 non-foil trials × 149 participants), df = degrees of freedom for χ^2 tests, *** $p < .001$. [a] The χ^2 tests were conducted with the next most complex model (i.e., simple was compared with null, full was compared with simple).

Table 3
Coefficient estimates and analysis of variance for fixed effects in linear mixed effects model predicting expectation match ratings.

Predictor	Estimate	SE	df	Wald F	p
Simple model (ignoring local statistics)					
Cadence	−1.36	0.09	182.2	207.00	<0.001***
Item order	0.01	<0.01	231.7	22.40	<0.001***
Cadence × item order	<0.01	<0.01	193.1	0.19	0.661
Full model (incorporating local statistics)					
Cadence	−1.31	0.10	187.3	188.13	<0.001***
Item order	−0.01	0.01	1188.8	1.62	0.203
Cadence × item order	−0.03	0.01	1624.8	23.44	<0.001***
Item order × DC proportion	0.03	0.01	1128.8	8.63	<0.003**
Cadence × item order × DC proportion	0.05	0.01	1415.4	24.53	<0.001***

Note. $N = 5960$ (40 non-foil trials × 149 participants). SE = standard error, df = denominator degrees of freedom for Type III Wald F tests (which were determined with the Satterthwaite approximation). The predictor *cadence* was coded as 1 = DC, −1 = PAC. *** $p < .001$.

over the trials for *DC proportion* = 0.2 but increase once *DC proportion* reaches 0.5. Meanwhile, PAC ratings increase over the trials for *DC proportion* = 0.2 but are almost constant for *DC proportion* = 0.8.

4.3.3. Model comparison

Table 4 contains the in-sample model fits and model comparisons for the null, simple, and full EM models. The marginal R^2 of the simple and full EM models suggests that the fixed effects and interactions explain a considerable amount of variance on their own (as compared with the CC models); however, the R^2 values for the simple and full models themselves are once again very similar to each other.

The null model's conditional R^2 (proportion of variance explained by fixed and random effects) was reasonably high, but the conditional R^2 values for the simple and full models were considerably higher than those for the null model, suggesting that once again the explanatory power of the model improved with the addition of by-participant random slopes. However, the conditional R^2 values of the simple and full models are comparable.

Unsurprisingly, the simple model explained significantly more variance than the null model. Once again, despite their seemingly similar model fits, the full model also explained significantly more variance than the simple model according to the χ^2 test. Additionally, the full model had a lower BIC than the simple model (which in turn had a lower BIC than the null model), meaning that the inclusion of predictors that capture the local statistics of the stimulus set improved the tradeoff between model fit and model complexity.

5. Discussion

Though expectation adaptation takes place in several auditory domains, prior research has not found evidence of an adaptation process for rare harmonic patterns in music. However, this lack of evidence arises from modelling approaches that do not account for the timecourse of expectation adaptation and the distribution of possible harmonic

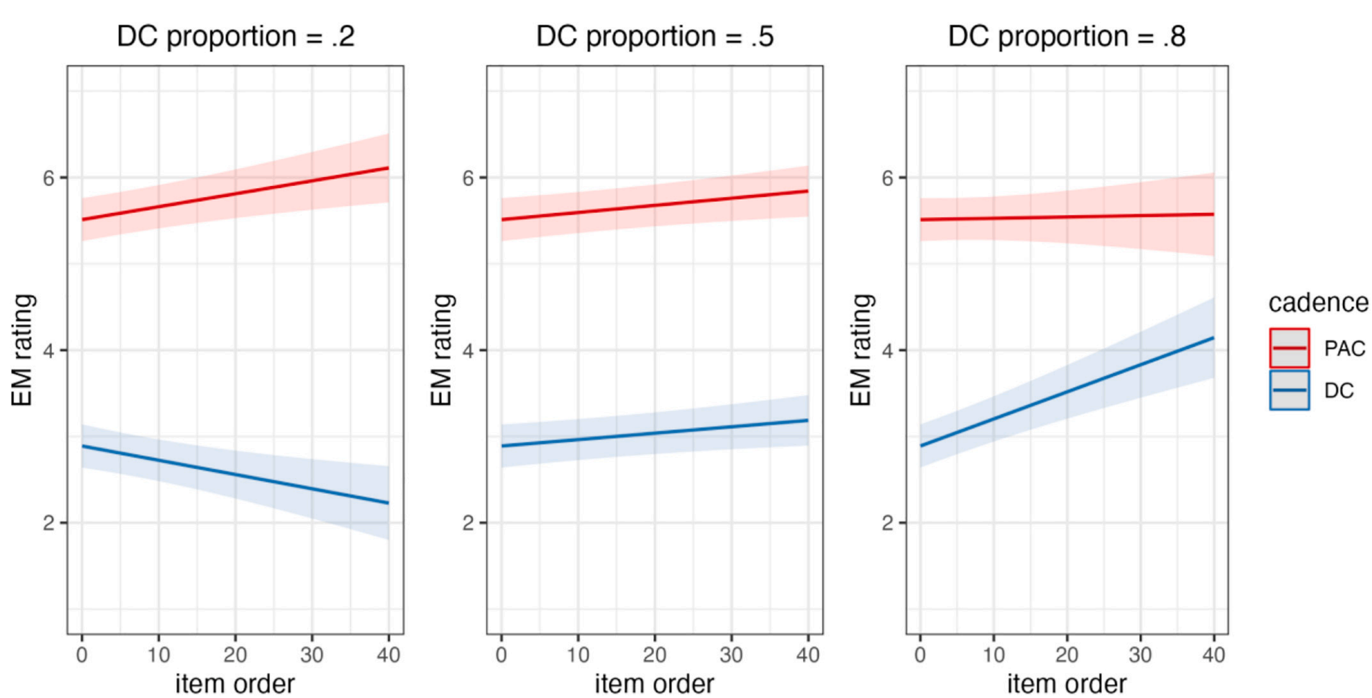

Fig. 4. Expectation match (EM) ratings estimated by the full linear mixed model that accounts for the local statistics of the stimulus set. Slopes are shown for the two cadence types at three different values of *DC proportion* (0.2, 0.5 and 0.8).

Table 4
In-sample model fits and model comparisons for null, simple and full models of expectation match ratings.

Model	Marginal R^2	Conditional R^2	Number of parameters	BIC	χ^2 value (df)[a]	p
Null	0	0.49	4	22,827	–	–
Simple	0.36	0.67	10	21,094	1784.55	<0.001***
Full	0.36	0.68	14	21,075	54.49	<0.001***

Note. $N = 5960$ (40 non-foil trials × 149 participants), df = degrees of freedom for χ^2 tests, *** $p < .001$. [a] The χ^2 tests were conducted with the next most complex model (i.e., simple was compared with null, full was compared with simple).

outcomes at a particular trial. In the present study, we measured participants' explicit ratings of how confident they were about which cadence would conclude a chord progression and how well the actual cadence matched their expectations. We directly modelled these ratings while accounting for the timecourse of the experiment and the probability of hearing an a priori rare cadence, namely a DC. We found that participants' completion confidence ratings changed over the course of the experiment: on average they increased when presented with more items, but this rate of change decreased as the probability of hearing a DC increased. Additionally, expectation match ratings were consistently lower for DCs than PACs, but ratings on average increased for both rare and common cadences as they progressed through the experiment. However, this effect was once more moderated by the local probability of hearing a DC at a particular trial: the higher this probability, the greater the increase in expectedness of a DC, and the smaller the increase in expectedness of a PAC. Together, these findings suggest that listeners can indeed adapt their expectations for rare harmonic patterns and provide support for the idea that Bayesian belief updating is a domain-general process that applies to music as well as language. In what follows, we discuss our key results, framing them in Bayesian terms and relating them to other findings in the literature.

5.1. Completion confidence ratings

We hypothesised that since participants heard far more DCs than would be predicted by corpus statistics, their confidence in their expectations for the "regular" cadence (i.e., PAC) would on average decrease over the course of the experiment, because the uncertainty of their posterior distribution would increase. Surprisingly, a model that accounted for the item order but did not account for the probability of hearing a DC at the current trial showed that CC ratings significantly *increased* by the end of the experiment, though this increase was small (only 0.01 per additional item). This result suggests one of two possible explanations. Firstly, if participants' confidence ratings index their long-term mental model of cadence likelihoods, then it is possible that they needed an even higher overall ratio of DCs to PACs to cause their confidence to decrease. This account assumes that listeners "explain away" the unexpected presence of DCs as due to noise (or a temporary deviation from random sampling) until there is greater evidence that the distribution of DCs has, in fact, increased. Secondly, since they needed time to get used to the experimental paradigm, their confidence rose once they had heard a few examples and became accustomed to the task. This account assumes that listeners require some minimal number of trials to appropriately "calibrate" their judgements in using the rating scale.

To test this first possibility, we ran a more complex model that also accounted for the local probability of hearing a DC by calculating the maximum likelihood estimation from all previous trials. This full model provided a better fit to the data than the simple model as per the BIC scores. We found that once the local probability of hearing a DC rose to 74%, the CC ratings began to decrease over the course of the experiment, suggesting that hearing a sufficiently high proportion of cadences that conflicted with our model of participants' prior knowledge did indeed affect their ability to predict how a musical phrase will end. However, the fact that ratings increased up to this probability suggests that there may well have been some adaptation to the task itself, independent of the statistics of the stimulus set.

While the marginal R^2 values (fixed effects only) for both the simple and full models were very low, the conditional R^2 values (fixed and random effects) were much higher. This suggests that there was considerable variation in how participants used the rating scales, but that accounting for this variation improved the fit. Interestingly, both types of R^2 were comparable between the simple and full model, but both the χ^2 test and BIC comparisons suggest that the full model had more explanatory power and was not overly complex compared to the simple model. On balance, we conclude that participants' confidence in how a phrase would be resolved was dependent on the changing probability of hearing a DC over the course of the experiment.

Confidence ratings for how phrases may end have not been employed extensively in music cognition studies. Related measures from other studies include Sears et al. (2019), who used ratings to measure the strength and specificity of participants' expectations for whether and how a phrase would continue (a task taken from Schmuckler, 1989). Sears, Caplin, and McAdams (2014) asked participants to rate the degree of completion of phrases and additionally asked participants how confident they were of their degree of completion rating. However, in our task, participants were told that the phrase would continue and only had to rate their confidence for the content of the completion. Moreover, the studies above did not investigate how confidence changed over time but instead measured the average confidence or expectancy strength and specificity for different phrase-final cadence types. However, our results do align with findings from Desender et al. (2021) who found that participants' confidence in the outcome of an event increased with additional evidence consistent with their initial belief and decreased when the evidence was inconsistent with their initial belief. Assuming that our participants were by default expecting a PAC (which is evidenced by the expectation match ratings), our results are also consistent with this account, though the threshold probability for this switch is surprisingly high.

5.2. Expectation match ratings

We hypothesised that EM ratings for DCs would be lower than PAC ratings, but that DC ratings would increase over the course of the study while PAC ratings would not increase. Our simple EM model (not accounting for the local probability of hearing a DC) found that DC ratings were indeed lower on average than PAC ratings. This result is consistent with evidence from both behavioural studies (Loui & Wessel, 2007; Sears et al., 2019; Tillmann & Bigand, 2010) and ERP studies (Janata, 1995; Koelsch et al., 2000, 2002; Koelsch et al., 2005; Steinbeis et al., 2006), as well as predictions from corpus analyses (de Clercq, 2015; Huron, 2006; Sears, Pearce, et al., 2018). They also parallel analogous situations in the language domain (Van Berkum et al., 2005).

However, while EM ratings increased significantly over the course of the experiment, there was no interaction in this simple model between the item order and cadence type, meaning that ratings for both PACs and DCs increased at approximately the same rate. As with the CC ratings, the lack of a significant interaction can be explained by the exclusion of the local probability of DCs from the model. Once local statistics were added, the model fit significantly improved, and the expected interaction between cadence and item order became significant, with DC ratings increasing faster than PAC ratings on average. This finding is consistent with a similar result in language from Fine et al. (2013), who found that participants read garden-path sentences that were resolved with a rare relative clause slower on average than sentences resolved with a common main verb, but their reading times sped up more for the relative clause resolutions over the course of the experiment.

Crucially, we observed a three-way interaction between the cadence, item order and local probability of hearing a DC. As the probability of hearing a DC increased, the EM ratings increased at a faster rate over the trials, and the PAC ratings in fact began to decrease slightly. This finding suggests that our participants were not only sensitive to the discrepancy between the global statistics of our stimuli and harmonic knowledge that they would have acquired implicitly through extensive exposure to Western music, but they also adapted in the short term to fluctuations in the local statistics of the stimulus set when rating how well the cadence matched their expectations. Our finding parallels results from Myslín and Levy (2016), who showed that changing the order of presentation of rare syntactic constructions so that they clustered together facilitated participants' processing, indexed by their faster reading times. However, Myslín and Levy only measured reading times after an exposure phase and averaged them across subsequent trials. Though they compensate for this to an extent by altering the amount of exposure that participants had, they do not capture the full timecourse of this adaptation as we do in our design.

The marginal R^2 values for both the simple and full models were moderate, indicating that the fixed effects alone explained a considerable amount of variation in the data. Once again, the conditional R^2 values (fixed and random effects) were considerably higher, implying that while there was considerable variation in how participants used the rating scales, accounting for this variation improved the fit. As was the case for the CC models, both types of R^2 were comparable between the simple and full EM models, but both the χ^2 test and BIC comparisons suggest that the full model had more explanatory power and was not overly complex compared to the simple model. Taken together with the equivalent findings for the CC models, we conclude that participants were sensitive to the changing probability of hearing a DC over the course of the experiment.

It is worthy of note that the participants' EM ratings for PACs remained high and did not change much throughout the experiment regardless of the proportion of DCs presented. This suggests that despite the observed adaptation for DCs, PACs did not in turn sound less expected for participants, perhaps because they are so strongly stored in long-term memory (Huron, 2006; Sears, Spitzer, Caplin, & McAdams, 2020; Tillmann & Bigand, 2010). However, the repetition of the a priori highly unexpected DC structure is not represented strongly in long-term memory and is thus acquired more readily through statistical learning without interfering with a listener's prior schematic knowledge.

5.3. Theoretical implications

Previous accounts of music perception have suggested that repetition priming has no effect on listeners' expectations for already familiar musical styles, as it is believed that schematic knowledge dominates the expectation formation process (Bigand et al., 2005; Tillmann & Bigand, 2010). These accounts are agnostic as to whether this knowledge is acquired through statistical learning (Huron, 2006; Pearce, 2018; Sears et al., 2019) or relates to the acoustic similarity of adjacent chords (Bharucha & Stoeckig, 1986; Tekman & Bharucha, 1998). However, these accounts do not factor in the timecourse of expectation adaptation, relying on analyses that average measurements across trials. By using a trial-by-trial modelling approach, we provide evidence that listeners use short-term statistical learning to adapt to the statistics of the stimulus set. Though short-term statistical learning helps listeners acquire the structure of unfamiliar musical scales (Loui et al., 2010) and artificial harmonic grammars (Jonaitis & Saffran, 2009), we show that it is also used when adapting to unusual but plausible cadences within a broadly familiar musical style.

More generally, we suggest that the musical expectation adaptation observed in our study is an instance of Bayesian inference, a domain-general process that explains a wide range of human behaviours (Chater, Oaksford, Hahn, & Heit, 2010). Our models incorporated predictors that accounted for how an ideal learner might iteratively update their priors based on the observed statistics of the musical style. These models provided a better fit to the data than models that only accounted for the timecourse of adaptation, suggesting that participants do indeed update their beliefs in real time when reporting their expectations for upcoming musical events.

Computational models of music perception such as IDyOM also operationalise the Bayesian inference process as combining long-term musical knowledge with the short-term acquisition of statistical contingencies in musical structure to generate expectations for upcoming musical events. However, while IDyOM is a flexible tool for modelling musical structure and captures higher-order dependencies between stimuli that our model cannot, it is not currently optimised for representing harmony. Since we were only using two cadence types in our study, and since the IDyOM configuration that we tried for the purposes of stimulus selection overestimated the difference in surprisal between PACs and DCs, we opted to use a simpler modelling approach. Once procedures for modelling harmony with IDyOM are more established, future work could examine expectation adaptation for a wider range of harmonic patterns.

Though we limited our participants to residents of the United Kingdom and the United States, we speculate that our results would be robust for participants primarily enculturated in musical styles that feature similarly referential chords (i.e., major and minor triads as opposed to diminished chords and other dissonant harmonies) and cadences akin to those in Western classical music. However, they may not extend to listeners who have extensive experience with styles that conflict with the cadential logic of Western classical music. For instance, jazz musicians (who frequently play music with complex harmonic logic and dissonant chords) have been shown to process cadences faster than musicians trained in other styles, and crucially do not exhibit a difference in reaction times for unexpected or deceptive cadences compared to authentic cadences (Adams, 2022). Thus, they may not exhibit much of an adaptation effect on the current task given that their reaction times across trials were already comparable for different cadences. However, they may not apply their expectations for jazz harmony to our classical stimuli. In any case, it appears that the level of expertise and enculturation in a musical style are important components in the account of expectation adaptation. Future work could examine this more explicitly.

6. Conclusion and future directions

In conclusion, we have shown that repetition priming can influence listeners' musical expectations in the short term. Moreover, we demonstrate that models accounting for how the statistical structure of the environment changes in real time provide a better fit to explicit ratings of confidence and expectedness, suggesting that item order matters in repetition priming. In future work, we plan to address expectation adaptation in listening contexts that are less homogenous than the present case in which all exemplars came from the same genre of Bach chorales. For example, hierarchical Bayesian models allow for more complex priors in which observations are placed into categories, each of which has their own set of expectations. An ideal learner should adopt such a hierarchical model so that expectation adaptation only applies to the relevant category and is not applied uniformly across musical genres. We would also like to explore whether there are neural signatures of short-term expectation adaptation, and whether thematic organisation in musical structure predicts the faster adaptation to rare events that we observed in participants hearing more DCs earlier on in the experiment.

CRediT authorship contribution statement

Aditya Chander: Conceptualization, Methodology, Investigation, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. **Richard N. Aslin:** Conceptualization, Methodology, Supervision, Writing – review & editing, Funding acquisition.

Declaration of Competing Interest

None.

Data availability

Our stimuli, data and analyses are available at https://osf.io/r8mzf/.

Acknowledgements

This work was supported by an NIH research grant HD-037082 to Richard N. Aslin (Elissa Newport, PI).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cognition.2023.105601.

References

Adams, L. (2022). Surprised by sound: Investigating expert musicians' and general listeners' reactions to harmonic surprises in jazz and popular music. In *Joint SMI/ICTM Ireland Postgraduate Conference 2022*. Dublin, Ireland: Dublin City University. https://www.researchgate.net/publication/358131018_Surprised_by_Sound_Investigating_expert_musicians'_and_general_listeners'_reactions_to_harmonic_surprises_in_jazz_and_popular_music.

Agres, K., Abdallah, S., & Pearce, M. (2018). Information-theoretic properties of auditory sequences dynamically influence expectation and memory. *Cognitive Science, 42*(1), 43–76. https://doi.org/10.1111/cogs.12477

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition, 73*(3), 247–264. https://doi.org/10.1016/S0010-0277(99)00059-1

Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods, 52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x

Aslin, R. N. (2017). Statistical learning: A powerful mechanism that operates by mere exposure. *WIREs Cognitive Science, 8*(e1373), 1–7. https://doi.org/10.1002/wcs.1373

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9*(4), 321–324. https://doi.org/10.1111/1467-9280.00063

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*, 1–48. https://doi.org/10.18637/jss.v067.i01

Besson, M., & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 1278–1296. https://doi.org/10.1037/0096-1523.21.6.1278

Bharucha, J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception, 5*(1), 1–30. https://doi.org/10.2307/40285384

Bharucha, J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 12*(4), 403–410. https://doi.org/10.1037/0096-1523.12.4.403

Bigand, E. (2003). More about the musical expertise of musically untrained listeners. *Annals of the New York Academy of Sciences, 999*(1), 304–312. https://doi.org/10.1196/annals.1284.041

Bigand, E., Delbé, C., Poulin-Charronnat, B., Leman, M., & Tillmann, B. (2014). Empirical evidence for musical syntax processing? Computer simulations reveal the contribution of auditory short-term memory. *Frontiers in Systems Neuroscience, 8*. https://www.frontiersin.org/articles/10.3389/fnsys.2014.00094.

Bigand, E., & Pineau, M. (1997). Global context effects on musical expectancy. *Perception & Psychophysics, 59*(7), 1098–1107. https://doi.org/10.3758/BF03205524

Bigand, E., Poulin, B., Tillmann, B., Madurell, F., & D'Adamo, D. A. (2003). Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance, 29*(1), 159–171. https://doi.org/10.1037/0096-1523.29.1.159

Bigand, E., Tillmann, B., Poulin-Charronnat, B., & Manderlier, D. (2005). Repetition priming: Is music special? *The Quarterly Journal of Experimental Psychology Section A, 58*(8), 1347–1375. https://doi.org/10.1080/02724980443000601

Bubic, A., Von Cramon, D. Y., & Schubotz, R. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience, 4*. https://www.frontiersin.org/articles/10.3389/fnhum.2010.00025.

Caplin, W. E. (1998). *Classical form: A theory of formal functions for the instrumental music of Haydn, Mozart, and Beethoven*. USA: Oxford University Press.

Caplin, W. E. (2004). The classical cadence: Conceptions and misconceptions. *Journal of the American Musicological Society, 57*(1), 51–118. https://doi.org/10.1525/jams.2004.57.1.51

Chater, N., Oaksford, M., Hahn, U., & Heit, E. (2010). Bayesian models of cognition. *WIREs Cognitive Science, 1*(6), 811–823. https://doi.org/10.1002/wcs.79

Cheung, V. K. M., Harrison, P., Koelsch, S., Pearce, M., Friederici, A. D., & Meyer, L. (2020). Distinct roles of cognitive and sensory information in musical expectancy. *PsyArXiv.*. https://doi.org/10.31234/osf.io/z76hg

de Clercq, T. (2015). A model for scale-degree reinterpretation: Melodic structure, modulation, and cadence choice in the chorale harmonizations of J. S. Bach. *Empirical Musicology Review, 10*(3). https://doi.org/10.18061/emr.v10i3.4334. Article 3.

Collins, T., Tillmann, B., Barrett, F., Delbé, C., & Janata, P. (2014). A combined model of sensory and cognitive representations underlying tonal expectations in music: From audio signals to behavior. *Psychological Review, 121*(1), 33–65. https://doi.org/10.1037/a0034695

Creel, S. C., Newport, E. L., & Aslin, R. N. (2004). Distant melodies: Statistical learning of nonadjacent dependencies in tone sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1119–1130. https://doi.org/10.1037/0278-7393.30.5.1119

Cuthbert, M. S., & Ariza, C. (2010). music21: A toolkit for computer-aided musicology and symbolic music data. In *Proceedings of the 11th International Society for Music Information Retrieval Conference* (pp. 637–642).

Desender, K., Donner, T. H., & Verguts, T. (2021). Dynamic expressions of confidence within an evidence accumulation framework. *Cognition, 207*, Article 104522. https://doi.org/10.1016/j.cognition.2020.104522

Ellis, C. T., Skalaban, L. J., Yates, T. S., Bejjanki, V. R., Córdova, N. I., & Turk-Browne, N. B. (2021). Evidence of hippocampal learning in human infants. *Current Biology, 31*(15), 3358–3364.e4. https://doi.org/10.1016/j.cub.2021.04.072

Federmeier, K. D. (2007). Thinking ahead: The role and roots of prediction in language comprehension. *Psychophysiology, 44*(4), 491–505. https://doi.org/10.1111/j.1469-8986.2007.00531.x

Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PLoS One, 8*(10), Article e77661. https://doi.org/10.1371/journal.pone.0077661

Fiser, J., & Aslin, R. N. (2002). Statistical learning of higher-order temporal structure from visual shape sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 458–467. https://doi.org/10.1037/0278-7393.28.3.458

Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron, 47*(4), 495–501. https://doi.org/10.1016/j.neuron.2005.06.020

Gjerdingen, R. (2007). *Music in the galant style*. Oxford University Press.

Gold, B. P., Pearce, M. T., Mas-Herrero, E., Dagher, A., & Zatorre, R. J. (2019). Predictability and uncertainty in the pleasure of music: A reward for learning? *Journal of Neuroscience, 39*(47), 9397–9409.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience, 30*(1), 535–574. https://doi.org/10.1146/annurev.neuro.29.051605.113038

Graf Estes, K., Evans, J. L., Alibali, M. W., & Saffran, J. R. (2007). Can infants map meaning to newly segmented words?: Statistical segmentation and word learning. *Psychological Science, 18*(3), 254–260. https://doi.org/10.1111/j.1467-9280.2007.01885.x

Haith, M. M., Hazan, C., & Goodman, G. S. (1988). Expectation and anticipation of dynamic visual events by 3.5-month-old babies. *Child Development, 59*(2), 467–479. https://doi.org/10.2307/1130325

Hansen, N. C., & Pearce, M. T. (2014). Predictive uncertainty in auditory sequence processing. *Frontiers in Psychology, 5*. https://www.frontiersin.org/articles/10.3389/fpsyg.2014.01052.

Harrison, D. (2020). Cadence. In A. Rehding, & S. Rings (Eds.), *The Oxford handbook of critical concepts in music theory* (pp. 534–576). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780190454746.013.1.

Harrison, P. (2020). *Modelling the perception and composition of Western musical harmony*. Thesis. Queen Mary University of London https://qmro.qmul.ac.uk/xmlui/handle/123456789/68149.

Harrison, P., Bianco, R., Chait, M., & Pearce, M. (2020). PPM-decay: A computational model of auditory prediction with memory decay. *PLoS Computational Biology, 16*(11), Article e1008304. https://doi.org/10.1371/journal.pcbi.1008304

Harrison, P., & Pearce, M. (2018). *Dissociating sensory and cognitive theories of harmony perception through computational modeling*. Proceedings of ICMPC15/ESCOM10. International Conference on Music Perception and Cognition, Graz, Austria.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

Huron, D., & Margulis, E. H. (2010). Musical expectancy and thrills. In *Handbook of music and emotion: Theory, research, applications* (pp. 575–604). Oxford University Press.

Isbilen, E. S., McCauley, S. M., & Christiansen, M. H. (2022). Individual differences in artificial and natural language statistical learning. *Cognition, 225*, Article 105123. https://doi.org/10.1016/j.cognition.2022.105123

Isbilen, E. S., McCauley, S. M., Kidd, E., & Christiansen, M. H. (2017). Testing statistical learning implicitly: A novel chunk-based measure of statistical learning. In *Proceedings of the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)* (pp. 564–569). Austin, TX: Cognitive Science Society.

Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience, 7*(2), 153–164. https://doi.org/10.1162/jocn.1995.7.2.153

Jonaitis, E., & Saffran, J. (2009). Learning harmony: The role of serial statistics. *Cognitive Science, 33*(5), 951–968. https://doi.org/10.1111/j.1551-6709.2009.01036.x

Jones, M. R. (2018). *Time will tell: A theory of dynamic attending*. Oxford University Press.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Lang & Ling Compass, 2* (4), 647–670. https://doi.org/10.1111/j.1749-818X.2008.00072.x

Keiflin, R., & Janak, P. H. (2017). Error-driven learning: Dopamine signals more than value-based errors. *Current Biology, 27*(24), R1321–R1324. https://doi.org/10.1016/j.cub.2017.10.043

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of Cognitive Neuroscience, 12*(3), 520–541. https://doi.org/10.1162/089892900562183

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience, 17*(10), 1565–1577. https://doi.org/10.1162/089892905774597290

Koelsch, S., Schmidt, B.-H., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology, 39* (5), 657–663. https://doi.org/10.1017/S0048577202010508

Kok, P., Jehee, J. F. M., & de Lange, F. P. (2012). Less is more: Expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265–270. https://doi.org/10.1016/j.neuron.2012.04.034

Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature, 427*(6971). https://doi.org/10.1038/nature02169. Article 6971.

Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. Oxford University Press.

Krumhansl, C., & Kessler, E. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*(4), 334–368. https://doi.org/10.1037/0033-295X.89.4.334

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647. https://doi.org/10.1146/annurev.psych.093008.131123

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13). https://doi.org/10.18637/jss.v082.i13

Logan, G. D. (1990). Repetition priming and automaticity: Common underlying mechanisms? *Cognitive Psychology, 22*(1), 1–35. https://doi.org/10.1016/0010-0285(90)90002-L

Loui, P., & Wessel, D. (2007). Harmonic expectation and affect in Western music: Effects of attention and training. *Perception & Psychophysics, 69*(7), 1084–1092. https://doi.org/10.3758/BF03193946

Loui, P., Wessel, D. L., & Hudson Kam, C. L. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception, 27*(5), 377–388. https://doi.org/10.1525/mp.2010.27.5.377

MacDonald, M. C., Just, M. A., & Carpenter, P. A. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology, 24*(1), 56–98. https://doi.org/10.1016/0010-0285(92)90003-K

Margulis, E. H. (2005). A model of melodic expectation. *Music Perception, 22*(4), 663–714. https://doi.org/10.1525/mp.2005.22.4.663

Margulis, E. H. (2014). *On repeat: How music plays the mind*. OUP USA.

Meyer, L. B. (1957). Meaning in music and information theory. *The Journal of Aesthetics and Art Criticism, 15*(4), 412–424. https://doi.org/10.2307/427154

Morgan, E., Fogel, A., Nair, A., & Patel, A. (2019). Statistical learning and gestalt-like principles predict melodic expectations. *Cognition, 189*, 23–34. https://doi.org/10.1016/j.cognition.2018.12.015

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS One, 9*(2), Article e89642. https://doi.org/10.1371/journal.pone.0089642

Myslín, M., & Levy, R. (2016). Comprehension priming as rational expectation for repetition: Evidence from syntactic processing. *Cognition, 147*, 29–56. https://doi.org/10.1016/j.cognition.2015.10.021

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142. https://doi.org/10.1111/j.2041-210x.2012.00261.x

Neuwirth, M., & Bergé, P. (Eds.). (2015). *What is a cadence? Theoretical and analytical perspectives on cadences in the classical repertoire*. Leuven University Press. https://doi.org/10.2307/j.ctt14jxt45.

Newport, E. L., & Aslin, R. N. (2004). Learning at a distance I. statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48*(2), 127–162. https://doi.org/10.1016/S0010-0285(03)00128-2

Ohriner, M. (2013). Effects of temporal position on harmonic succession in the Bach chorale corpus. In J. Yust, J. Wild, & J. A. Burgoyne (Eds.), *Mathematics and computation in music* (pp. 167–176). Springer. https://doi.org/10.1007/978-3-642-39357-0_13.

Pearce, M. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition*. London: City University.

Pearce, M. (2018). Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. *Annals of the New York Academy of Sciences, 1423*(1), 378–395. https://doi.org/10.1111/nyas.13654

Poulin-Charronnat, B., Bigand, E., & Koelsch, S. (2006). Processing of musical syntax tonic versus subdominant: An event-related potential study. *Journal of Cognitive Neuroscience, 18*(9), 1545–1554. https://doi.org/10.1162/jocn.2006.18.9.1545

Quinn, I. (2010). Are pitch-class profiles really "key for key"? *Zeitschrift Der Gesellschaft Für Musiktheorie [Journal of the German-Speaking Society of Music Theory], 7*(2), 151–163. https://doi.org/10.31751/513

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science, 274*(5294), 1926–1928. https://doi.org/10.1126/science.274.5294.1926

Saffran, J. R., Johnson, E. K., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition, 70*(1), 27–52. https://doi.org/10.1016/S0010-0277(98)00075-4

Sauvé, S. A., & Pearce, M. T. (2019). Information-theoretic modeling of perceived musical complexity. *Music Perception, 37*(2), 165–178. https://doi.org/10.1525/mp.2019.37.2.165

Schmuckler, M. A. (1989). Expectation in music: Investigation of melodic and harmonic processes. *Music Perception, 7*(2), 109–149. https://doi.org/10.2307/40285454

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.

Sears, D., Caplin, W. E., & McAdams, S. (2014). Perceiving the classical cadence. *Music Perception, 31*(5), 397–417. https://doi.org/10.1525/mp.2014.31.5.397

Sears, D., Korzeniowski, F., & Widmer, G. (2018). Evaluating language models of tonal harmony. In *Proceedings of the 19th International Society for Music Information Retrieval Conference* (pp. 211–217). https://doi.org/10.5281/zenodo.1492385

Sears, D., Pearce, M., Caplin, W., & McAdams, S. (2018). Simulating melodic and harmonic expectations for tonal cadences using probabilistic models. *Journal of New Music Research, 47*(1), 29–52. https://doi.org/10.1080/09298215.2017.1367010

Sears, D., Pearce, M., Spitzer, J., Caplin, W., & McAdams, S. (2019). Expectations for tonal cadences: Sensory and cognitive priming effects. *Quarterly Journal of Experimental Psychology, 72*(6), 1422–1438. https://doi.org/10.1177/1747021818814472

Sears, D., Verbeten, J., & Percival, H. (2021). Intonation discrimination for tonal chord sequences in a priming paradigm: Effects of target predictability and musical expertise. *Auditory Perception & Cognition, 1–16*. https://doi.org/10.1080/25742442.2021.1972744

Sears, D. R. W., & Forrest, D. (2021). Triadic patterns across classical and popular music corpora: Stylistic conventions, or characteristic idioms? *Journal of Mathematics and Music, 15*(2), 140–153. https://doi.org/10.1080/17459737.2021.1925762

Sears, D. R. W., Spitzer, J., Caplin, W. E., & McAdams, S. (2020). Expecting the end: Continuous expectancy ratings for tonal cadences. *Psychology of Music, 48*(3), 358–375. https://doi.org/10.1177/0305735618803676

Sears, D. R. W., Verbeten, J. E., & Percival, H. M. (2023). Does order matter? Harmonic priming effects for scrambled tonal chord sequences. *Journal of Experimental Psychology. Human Perception and Performance*. https://doi.org/10.1037/xhp0001103

Slevc, L. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review, 16*(2), 374–381. https://doi.org/10.3758/16.2.374

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience, 18*(8), 1380–1393. https://doi.org/10.1162/jocn.2006.18.8.1380

Stephenson, K. (2002). *What to listen for in rock: A stylistic analysis*. Yale University Press.

Summerfield, C., & Egner, T. (2009). Expectation (and attention) in visual cognition. *Trends in Cognitive Sciences, 13*(9), 403–409. https://doi.org/10.1016/j.tics.2009.06.003

Tekman, H. G., & Bharucha, J. J. (1998). Implicit knowledge versus psychoacoustic similarity in priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 252–260. https://doi.org/10.1037/0096-1523.24.1.252

Temperley, D. (2011). The cadential IV in rock. *Music Theory Online, 17*(1). https://mtosmt.org/issues/mto.11.17.1/mto.11.17.1.temperley.html.

Tillmann, B., & Bigand, E. (2010). Musical structure processing after repeated listening: Schematic expectations resist veridical expectations. *Musicae Scientiae, 14*(2_suppl), 33–47. https://doi.org/10.1177/10298649100140S204

Tillmann, B., Janata, P., Birk, J., & Bharucha, J. J. (2003). The costs and benefits of tonal centers for chord processing. *Journal of Experimental Psychology: Human Perception and Performance, 29*(2), 470–482. https://doi.org/10.1037/0096-1523.29.2.470

Van Berkum, J. J. A., Brown, C. M., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 443–467. https://doi.org/10.1037/0278-7393.31.3.443

Västfjäll, D. (2010). Indirect perceptual, cognitive, and behavioural measures. In P. N. Juslin (Ed.), *Handbook of music and emotion: Theory, research, applications* (p. 0). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199230143.003.0010.

Wall, L., Lieck, R., Neuwirth, M., & Rohrmeier, M. (2020). The impact of voice leading and harmony on musical expectancy. *Scientific Reports, 10*(1). https://doi.org/10.1038/s41598-020-61645-4. Article 1.

Waters, G. S., & Caplan, D. (1996). Processing resource capacity and the comprehension of garden path sentences. *Memory & Cognition, 24*(3), 342–355. https://doi.org/10.3758/BF03213298

Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology, 8*(2), 227–233. https://doi.org/10.1016/S0959-4388(98)80144-X