**The effect of explicit learning on prediction and uncertainty in serial music**

Sarah A. Sauvé, Alex Cho, Joe Argentino, Benjamin R. Zendel

Predictive coding is a compelling theory of brain function that has gained much interest in recent years (Clark, 2013b, 2013a; Friston & Kiebel, 2011). According to this framework, our brains deal in prediction and *prediction error*. Models of the world generate predictions and incoming sensory information generate prediction errors; our brain then works to explain away these errors by updating its models to make better predictions. These predictions are further modulated by uncertainty; given a context, a prediction may be a total guess or it may be the only perceived option. Given its highly structured nature, music provides an interesting context in which to investigate the predictive coding theory. The predictive coding framework has been applied to some aspects of music perception (Koelsch, Vuust, & Friston, 2018; Rohrmeier & Koelsch, 2012) including the perception of groove (Vuust, Ostergaard, Pallesen, Bailey, & Roepstorff, 2009; Vuust & Witek, 2014) and auditory scene analysis (Bendixen, 2014).

Beyond these specific applications, the concept of predictability and expectation is not new to music science (Meyer, 1956). It has been applied to a variety of music perception topics including melody perception (Margulis, 2005; Narmour, 1990, 1992; Pearce & Wiggins, 2006; Schellenberg, 1997; Temperley, 2008), phrase segmentation (Pearce, Müllensiefen, & Wiggins, 2010), musical emotion (Egermann, Pearce, Wiggins, & McAdams, 2013; Sauvé, Sayed, Dean, & Pearce, 2018), perceived complexity (Eerola, 2016; Sauvé & Pearce, in press) and auditory scene analysis (French-St. George & Bregman, 1989; Schröger et al., 2014). One promising computational model of musical expectations is IDyOM, or information dynamics of music (Pearce, 2005, 2018). IDyOM is a variable-order Markov model (Begleiter, El-Yaniv, & Yona, 2004; Bunton, 1997) which uses a multiple-viewpoint framework (Conklin & Witten, 1995), allowing it to combine models of different representations of the musical surface. IDyOM uses statistical learning to build models of the structural regularities in music and then uses these models to generate probabilistic predictions for a musical event based on the preceding context. Given a context, IDyOM estimates the probability of different continuations of the context based on how often they have appeared in similar contexts in its previous experience of music. IDyOM’s predictions combine probabilities derived from a long-term model trained on a large corpus, reflecting schematic learning of structure through long-term exposure to a musical style, and a short-term model, trained incrementally on the current piece of music, reflecting learning of local learning of motivic structure internal to a piece of music. IDyOM can generate probabilistic predictions for the pitch and timing of a musical note in a melodic context and the next chord in a harmonic sequence. IDyOM can also generate the entropy associated with these predictions. Entropy measures the uncertainty of each prediction, where maximum entropy occurs when all possible events are equiprobable.

IDyOM has been shown to accurately predict Western listeners’ pitch expectations in behavioural, physiological and EEG studies (e.g., Egermann et al., 2013; Hansen & Pearce, 2014; Omigie, Pearce, & Stewart, 2012; Omigie, Pearce, Williamson, & Stewart, 2013; Pearce, 2005; Pearce, Ruiz, Kapasi, Wiggins, & Bhattacharya, 2010; Quiroga-Martinez et al., 2019). In many circumstances, IDyOM provides a more accurate model of listeners’ pitch expectations than static rule-based models (e.g., Narmour, 1990; Schellenberg, 1997).

**Tonality and Atonality**

The vast majority of Western classical music is *tonal*. Tonal music is based on the concept of key, where the tonic, or first note, of that key is the most important note. The other pitches of the key have a very specific pattern of importance which forms the tonal hierarchy (Krumhansl & Shepard, 1979). Atonal music is a rejection of this tonal hierarchy, a movement beginning in the 19th century. Serial music is a specific style of post-tonal music invented by Arnold Schoenberg (Schoenberg, 1975) whose basic unit is the *tone row* and whose goal is to avoid tonal centers. The tone row consists of all twelve notes of the chromatic scale. This row can be transformed in several ways: it can be reversed, inverted or reversed and inverted. The row and its transformations make up the basic melodic and harmonic material of the music.

Research on the perception of atonal music is also a minority (Ball, 2011; Imberty, 1993). Without tonal structure to guide listeners, alternative sources of structure have been explored (Dibben, 1994, 1999; Lerdahl, 1989) as well as the lingering role of the tonal hierarchy (Krumhansl, Sandell, & Sergeant, 1987). Krumhansl et al. (1987) applied the probe tone method (Krumhansl & Shepard, 1979) to tone rows from Schoenberg’s Wind Quintet and String Quartet No. 4. Large individual differences were observed; however, two general patterns could be observed. First, listeners with little to no knowledge of serial music presented traces of the tonal hierarchy pattern in their responses. In contrast, listeners with knowledge serial music presented a pattern of expectations that explicitly avoided a tonal center. In other words, they had a model of expectations specific to serial music. This type of stylistic compartmentalization has also been observed between classical and jazz musicians (Hansen, Vuust, & Pearce, 2016).

**Learning Atonality**

This study applies the predictive coding framework to serial music, posing two core questions. How does explicitly learning about serial music change one’s perception of serial music in terms of expectancy and uncertainty? As a computational model of musical expectancy, can IDyOM model the perception of serial music as well as it does tonal music? What aspects of the musical surface best model human perception?

In order to answer these questions, expectancy and certainty ratings for serial phrase excerpts will be collected from undergraduate music students before and after they take a post-tonal theory course, which includes the study of serial music. Though some exposure and familiarity with serial music is possible, as a school steeped in the classical tradition, we expect most students will have minimal exposure. As such, initially their predictions will likely be poor and their uncertainty high (H1). However, with exposure to serial music and explicit knowledge of the style, we expect predictions to remain poor but uncertainty to lower (H2). Koelsch, Vuust & Friston (2019) call this *expected uncertainty*.

IDyOM will perform a similar experiment, generating information content (inverse probability) and entropy for the same serial phrases as the human participants with two different long-term models. The first will be trained on tonal music alone, mirroring human participants’ exposure before their course. The second will be trained on the same tonal music corpus, plus the serial phrases from the before portion of the behavioural study, mirroring the students’ exposure following the course. These models will be generated for pitch using the linked pitch and pitch interval viewpoints, as we hypothesize that this combination will model human perception best (H3). Furthermore, due to the purely statistical nature of the computational model and the statistical properties of serial music, we expect that the computational model that has been exposed to serial music will predict serial music more accurately and with better certainty than humans that have learned about serial music (H4).

Caveat: focus on pitch alone, where there’s evidence that atonal music has different hierarchical cues.

**Materials and methods**

**Participants**

34 third year students of Memorial University of Newfoundland’s School of Music and enrolled in the Materials and Techniques of Post-Tonal Music course participated in this study and provided written informed consent in accordance with the Interdisciplinary Committee on Ethics in Human Research at Memorial University of Newfoundland. Their mean age is XX (SD = XX), mean musical sophistication according to Goldsmith’s Musical Sophistication Index’s musical training subscale is XX (SD = XX) and exposure to atonal music according to our own devised scale (out of 35) is XX (SD = XX). All participants received a small cash honorarium for their participation.

**Stimuli**

Sixteen serial music phrases were selected for this study, eight for the Before portion and eight for the After portion. Table 1 presents the details of these excerpts; MIDI files can be found on the project’s OSF page[[1]](#footnote-1). Each phrase contains part or all of a tone row and is a fully formed musical phrase. Phrases were divided into Before or After to achieve an equal number of note events in each phase; both contain 105 note events. Phrases included in the After portion were not covered in class, while phrases included in the Before portion could be.

**Table 1**. Summary of stimuli

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Composer** | **Title of work** | **Measures** | **Instrument** | **Length** | **Time** |
| Luigi Dallapiccola | "Quartina," No. 11 from *Quaderno musicale di Annalibera* | 1-5 | Piano | 14 | Before |
| 6-9 | Piano | 13 | After |
| Arnold Schoenberg | Variations for Orchestra, Op. 31 | 51-57 | 1st violins | 13 | After |
| 34-38 | Cellos | 13 | Before |
| 46-50 | Cellos | 13 | After |
| 39-45 | Cellos | 14 | Before |
| Anton Webern | "Wie bin ich froh," No. 1 from *Drei Lieder*, Op. 25 | 3-5 | Voice | 16 | After |
| 6-8 | Voice | 12 | Before |
| "Des Herzens Purpurvogel fliegt durch Nacht," No. 2 from *Drei Lieder*, Op. 25 | 3-8 | Voice | 11 | Before |
| 10-15 | Voice | 11 | After |
| "Sterne, Ihr silbernen Bienen der Nacht," No. 3 from *Drei Lieder*, Op. 25 | 29-36 | Voice | 15 | After |
| 37-45 | Voice | 16 | Before |
| String Quartet, Op. 28, Mvmt. 2 | 1-7 | 1st violin | 12 | After |
| 8-14 | 1st violin | 12 | Before |
| 2-10 | 2nd violin | 13 | Before |
| 11-18 | 2nd violin | 12 | After |

**Notes**. Length is given in number of note events.

**Procedure**

After providing informed consent, participants were asked to fill out the Gold-MSI musical training sub-scale and exposure to atonal music questionnaires; both can be found on the project’s OSF page. They were then moved to a double-walled, electrically shielded sound-attenuating booth for the auditory task, presented through XXX headphones using e-Prime 3; standalone implementation application available on the project’s OSF page. Figure 1 illustrates the procedure of a single trial: each phrase was presented in fragments, with each subsequent fragment containing one more pitch than the last. After each fragment, the participant answered the corresponding question with a rating on a scale of 1 to 7. Fragments were presented at a rate of 100bpm and at a comfortable volume. The trial was complete once the entire phrase had been presented and each pitch rated. For half the phrases, participants rated their predictions and sense of closure, answering the following two questions: “*How surprising was the last note?*” and “*How well did the last note finish the phrase?*”. For the other half of the phrases, participants rated their precision, or sense of certainty by answering the following question: “*How certain of you of what note will come next?*”. Two practice trials were presented, one for each type of question, followed by the eight phrases appropriate to the phase (Before/After); phrases were presented in random order. Participants were notified of which type of questions they would be answering for a particular trial before the trial began. The type of questions collected for each phrase was counterbalanced across participants.

|  |
| --- |
|  |
| **Figure 1**. Trial procedure: the phrase is presented in fragments, each with one pitch more than the last; after each fragment, the participant will answer the appropriate question with a rating from 1 to 7. |

**Analysis**

All analyses are implemented in R 3.3.2. Precision ratings are transformed to a reversed scale so that high certainty ratings correspond to low entropy.

**Confirmatory analysis**. Using the *lme4* package (Bates, Mächler, Bolker, & Walker, 2015), a mixed effects multiple linear regression model is fitted as per the study design with maximal random effects (Barr, Levy, Scheepers, & Tily, 2013). Two models are fit: one for prediction and one for precision (certainty ratings for humans), where these are the predicted values. Both models include *Time* (Before, After) and *Type* (Human, IDyOM) as fixed effects, with interactions modelled. Random effects include random intercepts on participants and random slopes on phrase. The categorical variables are factors, where each level is compared to a base level. These base levels are Before for Time and Human for Type. Models are evaluated using Pearson’s correlation between the model’s predictions and the data along with the correlation’s 95% CIs. Variance explained by each model is tested by calculating the coefficient of determination R2. Statistical significance of each model is tested by a likelihood-ratio test between a null model (intercept and random effects) and the maximally fitted model. Statistical significance of each individual factor level for a given predictor is evaluated using 95% CIs, where an interval not including zero indicates a significant predictor. IDyOM information content and entropy are calculated using the linked pitch and pitch interval viewpoints for this omnibus analysis as well as for analyses testing Hypotheses 1, 2 and 4.

Alpha is set to .05 for the following tests. Both p-value and effect size are reported. In order to test H1, one-sided t-tests are conducted, where µ = 5 and the alternative is greater for prediction ratings and less for precision ratings. H2 is tested using paired t-tests comparing before and after prediction and precision ratings. The prediction test is one-sided, where the alternative is greater, while the precision test is two-sided. Equivalence tests, implemented using the TOSTER package (Lakens, 2017), are also conducted with a minimum upper for prediction and upper and lower for precision equivalence bound of 0.5 raw units, corresponding to half a point on the rating scale. This equivalence bound is selected loosely based on previous work using expectancy ratings and related to IDyOM (Egermann et al., 2013). H3 is tested using Pearson’s correlation coefficient, calculated between human ratings of prediction and IDyOM information content produced by each combination of musical properties (viewpoints). The same is done for human ratings of precision and IDyOM entropy. These tests are performed on pooled Before and After data and on Before and After data separately. Finally, H4 is tested using between-samples t-tests to compare human prediction ratings to information content and human precision ratings to entropy, all from the After phase. Both t-tests are one-sided, where the alternative is less, and all values are transformed to z-scores for accurate comparison between human ratings and IDyOM output.

**Exploratory analysis.** Though it is commonly used, a concern with the paradigm employed is the confounding of predictability ratings with the perception of closure, where pitches that are perceived as poor endings to a phrase are also considered more surprising. In order to measure the potential effect of this confound, closure ratings were collected alongside ratings of expectedness. First, the degree of correlation between expectancy and closure ratings is calculated, collapsed across Time. Second, linear mixed effects models with and without closure ratings as a covariate are compared. Variance explained by closure ratings offer an approximation of how much the percept of closure influences the rating of surprise/expectedness.

The compartmentalization of style is also an interesting phenomenon to explore here. The IDyOM configuration used in this study employs an LTM trained on either only tonal music (Before) or a combination of tonal and atonal music (After) in order to approximate the listeners’ musical exposure. However, given evidence of the compartmentalization of musical style, or the ability to generate different predictions based on the style listened to (e.g. Hansen, Vuust, & Pearce, 2016), the information content and entropy produced by an LTM trained only on atonal music to an LTM trained on both tonal and atonal music are compared. In order to evaluate the degree of compartmentalization between styles, the Pearson correlation between information content and entropy of the atonal-only and tonal+atonal trained LTM model and human ratings are compared.

1. <https://osf.io/kqzg8/> [↑](#footnote-ref-1)