This paper uses *n*-gram transition probabilities to investigate the interconnection of bass-line and top-line behaviors. The author creates a matrix of *n*-gram probabilities from *n*=0 to *n*=5 for both top- and bottom-voice data, converts the transitions to entropy measures, identifies the principle components within this distribution, and studies how those components vary between voices and between the major and minor modes. They also correlate the entropy matrixes of each voice within each piece to one another to see whether they correlate. The author concludes that top and bottom lines are somewhat connected, but also somewhat different.

Overall, the paper is well designed, has some interesting points, and is basically well written. I think that another round of revisions is needed, though, before the piece is publishable.

I have four main points on this topic. First: I think the author needs to motivate the move between transition probabilities and entropy. They are obviously extremely connected, and as I see it, the author is basically using entropy to create a weighted average of probabilities for each transition within a piece. That is, instead of using a long vector representing every single moment (each token) within the piece, the piece is instead represented by a vector indexed by each transition type, with the probability weighted by the number of times it occurs in the piece, and converted to bits. This is all fine; but, some of the psychological motivations of the paper appear to rely solely on entropy, while one could argue the use of entropy in the paper is a just computational shorthand for moment-to-moment transitional events. Again, these representations are interconnected, but the author should motivate these conceptual moves more thoroughly.

Second, the author takes a pretty contentious approach to ways that computer models can be "verified." Temperley (2012) put this issue well when he describes three components of computational validation against a ground truth of human cognition: 1) Is the input what a human mind takes in, 2) does the computational engineering mimic cognitive processes, and 3) does the model's output align with human assessments or behavior. It seems as though the validations that the author discusses focus entirely on #3. There are surely reasons to be skeptical of claims about neurodynamics being represented by transition tables and entropy measures, and the author should acknowledge this.

In fact, I'm just not convinced that this paper is about cognition! It seems like it's actually about finding some interesting compositional tendency that are certainly present within a corpus and may or may not be salient to cognition. In other words, I also would encourage the author to rethink the article's opening gambit, motivate the investigation not by statistical learning but rather using research questions about compositional and stylistic tendencies, and then relate the findings to cognition in the general discussion section.

To my mind, the discussion page 32 could be far more interesting and fruitful as a frame for this article. I *loved* the discussion about locating musical style and creativity. I feel like the

whole article could be designed around this research question; even if not, I'd strongly encourage the author to foreshadow this finding in the introduction.

Third, the author's final conclusions are potentially under-baked. They write, "in summary, this study suggested that TP distributions and the entropies of the melody and bass line interact with and are partly independent of each other." Framing the findings in such a way –that the findings simply point to a muddy grey area of partial interactions– makes one wonder whether there are any take-aways from the paper at all! It's quite hard to make the argument the author wants to about the bass/soprano conditional knowledge: their results basically show that bass and soprano lines *somewhat* correlate. Therefore, an alternate explanation could simply be that humans' statistical knowledge of these lines is 100% derived from their pairing with some noise in compositional systems. A further alternate explanation is that our knowledge entirely voice independent, and the rules and norms of these voices just happen to compositionally overlap. Now, neither of these explanations seems musically intuitive– my point is that the author's current explanation is not without holes, and their results don't exclude other explanations.

On this front, a theorist might complain that deeper issue is the fact that there isn't any inter-voice tracking that's going on. If you knew the interval between the top and bottom voices, you'd probably be better able to predict when the top and bottom voices will act similarly and when they'll act differently. But, the author's point is that these are two lines of potentially independent information, and they are going to study the data with a basic theory-neutral naiveté in order to identify whether entropy is, in fact, a good way to make sense of these streams of data. The author, should, regardless deal with this potential criticism.

Fourth, I'm concerned about the lack of dialogue with other work that addresses similar research questions. There's no reference to work by Ian Quinn, David Temperley, David Sears, Panos Mavromatis or Chris White who do work with transition probabilities, style, and key. The work coming from the SIMSAA and ELVIS projects from McGill's CIRMMT lab explicitly deals with this (Peter Schubert and Ichiro Fujinaga would be two researchers the author might look at). Additionally, representations usually involve pitch-event-oriented *n*-grams, where each object in the representation is a note or chord. Here, the author uses intervals, a choice with it completely fine– but, given that it is a departure from the norms, it warrants a few motivating words. (Note, though, that Quinn and Mavromatis (2011) use this very representation, so it is not without precedent in the literature.) (To be clear: if this paper ends up being resubmitted, I would hope to see a good dozen [if not more] citations discussed surrounding probabilistic approaches to melodic and bass motions.)

Finally, a few nit picks. Language use is pretty good, but not quite up to a publishable standard. At some point, some editor will need to go through and do some fine-tuned editing. For instance, there are articles before some nouns that don't idiomatically use articles. Sentences like, "the modelling approaches partially outperform experimental results under conditions that are impossible to replicate in an experimental approach" also made me turn my head– I think

the author means that computer models are better at performing certain tasks than behavioral experiments, but this isn't exactly what "overperform" means. In the sentence, "Using the distributions of TPs (information content) in each melody, bass, major, and minor of each piece of music…", I just don't know what *major and minor* mean. Finally, the word "tonality" appears to mean major or minor mode here; I'd encourage author to offer an explicit definition of that word (since it means different things in different scholarly discourses) or use the word *tonal mode,* which to me specifically means major or minor. (Some other specifics: on page 11, "in component 2" repeats; on page 17 you want "Bass" not "Base"; the bottom of page 34 has a wayward semicolon.)

I also think that the authors' equations could be better explained. I'm not sure what "(bits)" is doing in each, since Log(2) already indicates that the units will be bits. Also, I'm still not totally clear on what Equation 3 is doing. First I convinced myself that *i* and *j* are adjacent time points, but then convinced myself that we are summing over all values i in I and all values j in J, which I suppose means we're looking at all pairs of chords. But then, we're relying on the same vocabulary of chords (i.e., I=J). This is all to say that I need a little more hand holding for this equation.

Temperley, David. 2012. "Computational Models of Music Cognition." In Diana Deutsch (Ed.), *The Psychology of Music*, (3rd edition), 327-368. Elsevier.

Quinn, Ian, and Panayotis Mavromatis. 2011. "Voice Leading and Harmonic Function in Two Chorale Corpora." In *Mathematics and Computation in Music*, ed. by Carlos Agon, 230-240. Springer.