Some things Fred and I

Jackendoff

(for LerdahlFest, March 2018)

I’m delighted to be able to take part in this splendid occasion honoring Fred on his retirement. Working with Fred for those many years we were writing GTTM was a formative experience for me. It had a lot to do with my becoming not just a linguist but a cognitive scientist, thinking way out of the box. And in the course of it, I learned a hell of a lot about music from Fred – he was the source of pretty much all the musical sophistication in the theory.

Here we are at our book party back in 1983, god, that was 35 years ago.

We had no idea at the time that our book would have such lasting power. There’s no question that its strength is thanks to the collaboration – if I may be a little immodest, it felt kind of like Watson and Crick or – well, I’m pretty sure Fred wouldn’t like this comparison – Gilbert and Sullivan. Francis Crick himself writes “It is ... extremely important not to be trapped by one’s mistaken ideas. The advantage of intellectual collaboration is that it helps jolt one out of false assumptions.” Our work together was like that – we continually challenged ourselves to articulate every little step until we were both satisfied with it, and also with how it fit into the bigger picture.

As many of you can attest – probably more than I can – Fred has had a lot more to say since GTTM, and his work has continued to be groundbreaking and inspirational, even when he protested that what he really wanted to be doing was composing. As for me, when the book was done, I felt like I didn’t have a lot more to say about music, at least not without Fred in the neighborhood any more, and for the most part I turned my attention to other topics.

Still, from time to time, and especially recently, work I’ve done in linguistics seems to have some relevance to music, and it’s raised issues that Fred and I never really touched on. So today I’d like to talk about some of those. I should warn you that this is far from a scholarly exposition. I’m just throwing out a few half-baked ideas, in the hope that someone might find them worth picking up on. And given that I’ve hardly kept up at all with what’s been going on in
the field of music cognition, I’d probably better apologize in advance if I’ve stumbled into territory that someone or even lots of people have already explored – especially people in this room.

The issues I want to mention fall roughly into four categories: first, narrative structure in music; second, motivic parallelism; third, memory for music; and fourth, expectation or prediction in the course of processing music.

First, an issue of musical structure. It’s become evident that if there is any entity in language for which there are parallels to music, it’s not the sentence, it’s not syntax. Rather, we might find closer parallels in narrative structure, the way we describe a sequence of complex events. This is a kind of organization bigger than sentences. Narrative often involves a sense of building tension followed by resolution, notions that we apply easily to musical structure – much more easily than we can apply syntactic notions like subject and predicate.

This parallel became more prominent in my own thinking through working with my then-student Neil Cohn, who was doing and still is doing brilliant work on the grammar of sequential visual images – wordless comic strips. In both theoretical and experimental work, he has shown that people structure a series of cartoon panels not just as one image after another, but filling out a narrative arc.

The basic components resemble Aristotle’s theory of dramatic structure: (a) an optional Establisher that sets the scene (say the boy holding the bowling ball), (b) an Initial, which shows the start of an action (he’s preparing to throwing the ball), (c) a Peak (the moment of actually throwing the ball), and (d) an optional Release (the outcome of the event and his reaction) – and there are other elaborations that add drama. What makes this especially interesting is that each of these components can itself consist of a smaller narrative arc, and this can continue recursively, so we can have highly structured complex events. For instance, here are two six-panel strips, with quite different embedded narrative structures.
These structures can be realized either through language or through wordless strips or combinations. Each medium offers its own affordances.

If we try to extend this notion of narrative to music, what structure expresses it? My first impulse, coming from the GTTM perspective, was that narrative structure should be localized in the prolongational reduction, which encodes patterns of tonal tension and release. But I was disabused of this idea when I happened to think about the passacaglia in the last movement of Brahms’ 4th symphony. Here’s how it starts, just to remind you.

![Sheet Music](image)

It consists of 31 variations on an 8-measure theme, totalling about 250 measures, followed by a coda of about 60 measures. The prolongational head of each variation is E minor, with the exception of three variations in E major, about halfway through. So the prolongational reduction of the whole movement is something like this:
In other words, nothing much really happens prolongationally until about 80 percent of the way through the piece. This is of course hugely at odds with the way you experience the piece, which is (at least for me) one of the most dramatic pieces in the literature.

It’s constantly building to climaxes and subsiding, sometimes suddenly, sometimes gradually, receding to points of stasis or quiet agitation, only to slowly – or suddenly – build again. The E major section and the variation before it (the flute solo) progress at half the speed and represent a sustained point of repose, only to be brutally interrupted by the repetition of the opening declamation. None of this is reflected in the prolongational reduction. Brahms does it all with melody, dynamics, and texture.

The same sorts of observations go for other variation movements. Another powerful example is the variation movement in Schubert’s *Death and the Maiden* quartet, which I find totally heartbreaking in the way it builds in intensity. But its prolongational reduction is basically just a repeated progression from g minor to G major. More generally, in a variation movement, interchanging variations doesn’t change the larger levels of the prolongational reduction, but it often changes the dramatic effect. And even more radically, of course, drum music can convey tension and release without any tonality at all.

What I’m getting from this is that tonal tension and relaxation are only one component of tension and relaxation overall. So the question is how this more general component of narrative or dramatic structure works, what its structure is, and how it interacts with prolongational reduction. Does it apply only to large forms, or is there a sort of Schenkerian fractal pattern, whereby the same patterns appear at large and small scales? And a more basic question is why we like narratives, whether in music, language, or comic strips? Well, I’m not going to try to answer these questions today – as I said, my job today is just to pose questions. But in a moment I’ll bring up some connections to other issues.

Another kind of structure has come to my attention in connection with a couple of different linguistic projects. It seems relevant for talking about motivic parallelism, a topic that Fred and I shied away from.

Consider this pair of little guys:
You might describe them by saying “They’re the SAME, EXCEPT they have different curls on their heads.” This relation, that two things are the same except for one feature that’s different, is found all over cognition. Actually, William James noticed it. He made the point that you don’t need language to appreciate this relationship. It would pop out even if I hadn’t described it to you.

James notices that this SAME-EXCEPT relationship occurs in every modality of perception. For instance, it turns up in language. Two strings rhyme if they’re the SAME, EXCEPT for the material before the main-stressed vowel: duck and truck, or debility and utility, or for an extreme case, embraceable you and irreplaceable you. My own interest in SAME-EXCEPT comes from work on ellipsis that I’ve been doing with Peter Culicover and work on morphology that I’ve been doing recently, in collaboration with Jenny Audring. For example, Jenny and I think SAME-EXCEPT is just the right way to talk about irregular verbs: sing and sang are the SAME, EXCEPT for the vowel and EXCEPT for the tense.

The SAME-EXCEPT relation doesn’t just say these two things are almost the same – it pinpoints and marks the specific places where they differ. Jenny and I think these relationships are explicit in mental representation, even if they’re unconscious.

To see their influence, think about the familiar Necker cube. It’s ambiguous.

So if you have two Necker cubes, they should be four ways ambiguous.

But they aren’t – they’re only two ways ambiguous (or at least it’s very hard to see them as four ways ambiguous), and they switch aspects together. This is because they’re linked in an unconscious SAME relation that says whatever structure one of them has, so does the other. That is, when you notice two things are the same, you’re adding information to your understanding.

SAME-EXCEPT turns out to be a perfect tool for talking about motivic parallelism. This plays a big role in determining grouping and meter, as we pointed out in GTTM, and as Davy Temperley picked up on in his 2001 book.
Here’s the beginning of ‘Happy Birthday.’

This can be described as “The second phrase is the same as the first, except the last two notes are a step higher.” This is a typical sort of motivic parallelism. Treating it in terms of same-except has the virtue of letting us attend to the differences as well as the similarities. It’s more precise than saying “The second phrase is pretty much the same as the first.”

There’s more going on, though. There’s this common story structure in language that makes use of the same-except relation: the story relates two things that are basically the same, followed by one that’s different. We might call this pattern sample + same-except + different. Examples are the three bears, the three little pigs, any of the many stories about the king and his three sons, or the three competitors for the princess, and also the jokes that start “Three guys walk into a bar....” In each case, the first two set up a pattern which the third breaks. We’ve internalized this pattern, so we notice when the formula is broken. I came across a little Yiddish story in which the mama fly tells her three baby flies three dangerous things to stay away from: honey, wine, and fire. The first baby fly falls for honey and gets killed; the second falls for wine and gets killed. The third is going to be wiser, right? No, he falls for fire and gets killed too.

The pattern that this story doesn’t follow, sample + same-except + different, shows up in music too. For instance it’s the principle behind the 12-bar blues, AAB.

At a smaller scale, the sample + same-except + different pattern is ubiquitous. In Happy Birthday, the first two groups are followed by a group that starts the same but diverges at the third note. The situation can be more complex.

In our old favorite, the Mozart G minor symphony, the little three-note motive is repeated and then followed by something that’s a bit different: the same AAB pattern. Then the second two-bar group has the same rhythm as the first, and its descents are the same except at different pitches. The whole 4-measure group is followed by another 4-measure group that’s the same but a step lower and differently harmonized. And that’s followed by something different, our sample
+ same-except + different pattern again, at a larger scale. The something different at measure 10 starts the same as the original 3-note motive, but it goes its own way, then is almost repeated but again goes off its own way in the second half of measure 13. The cadential chords in measures 16 to 19 compress the preceding 2 measures, again repeated, and in the bass is a motive that’s the same as the initial three notes but different in pitch. And these are followed in bar 19 and 20 by something that’s the same but further compressed rhythmically. And then the first group returns, and so on. All of this is pretty obvious.

I don’t know how to notate all these same-except relations, but they add up to a huge motivic density in the opening of this piece – it’s shot through with links based on same-except and especially sample + same-except + different.

Just for comparison, here’s the opening of Mozart’s clarinet concerto, where there’s much less in the way of motivic parallelism going on, compared even to the first four measures of the symphony. (Maybe the descent of a third in measure 1, followed by the speeded up descents in measure 3 count, but I think this is a pretty flimsy relation.)

And here’s the opening of the clarinet quintet, with a similar story. Maybe the deceptive cadences in measures 2 and 4 count as parallel, but again it’s a weak connection, and then of course the clarinet figuration in measure 8 is made up of little bits that are the same in contour but not in pitch – a characteristic use of Same-Except in sequences. On the other hand, the opening of this quintet is fairly dense harmonically, compared to either the symphony or the concerto – there are lots more branches in the harmonic parts of its prolongational tree.
* So one way in which passages of music can differ is in their density of motivic connections, and another is the density of their harmonic progressions.

At larger levels of structure, same-except relations are of course basic to variation forms, including our Brahms passacaglia, as well as to sonata form and just about every other form you can think of. The important thing to remember is that it’s not just the sameness that makes the music work, the excepts are just as important.

What I want to take from this is first, that motivic parallelism can be addressed in terms of the same-except relation, which is found everywhere in cognition; second, same-except can be fractal, applying from the smallest to the largest scales; third, an important part of our musical understanding is a network of explicit same-except relations within a piece – and across pieces as well.

That’s all I want to say about musical structure – again stressing that it’s more a sketch than a complete story.

Now I want to talk about the third issue, the nature of musical memory. How do you store pieces of music in your brain, and how does that impact on how you hear music?

In the literature that compares music and language, it’s often said (and Fred and I said it, I think) that music doesn’t have much of a counterpart of words, and therefore that it doesn’t make a lot of sense to talk about a lexicon in music. I’ll agree that music doesn’t have a counterpart of words, just as it doesn’t have a counterpart of sentences. But my thinking has changed radically when it comes to the lexicon, the “place” where words are stored – and it’s changed in such a way that a comparison with music shows far more promise. The basic idea is that the lexicon contains all kinds of things, and aside from words, music has counterparts of just about all of them.

So what’s in this new version of the lexicon? Well, first, you have to store idioms like *call the shots* and *pie in the sky* in your lexicon, because the meanings of their words don’t add up to their meaning as a whole. And they’re not just strings of words – they have ordinary syntactic structure. Syntactically, *call the shots* is an ordinary verb phrase that contains an ordinary noun phrase. And there are thousands of idioms. Second, the evidence from psycholinguistics is that
you also store lots of frequent clichés, whether they’re idiomatic or not, things like *I don’t think so, without further ado, there’s a sense in which ...*, and so on – again with internal syntactic structure.

Third, things that we’ve been accustomed to thinking of as *rules of grammar* are better thought of as templates or *schemata* that define a piece of linguistic structure, with slots that can be filled by other material. For instance, the phrase structure rule for a transitive verb phrase, VP \( \rightarrow V - NP \) is normally understood to say ‘expand a verb phrase as a verb followed by a noun phrase.’ You can recast this as a template or schema \([_{VP} V \_NP]\), which says ‘a verb phrase can consist of a verb followed by a noun phrase.’

This not only generates novel VPs like *hire a zookeeper*, it also supports or motivates the structure of stored VP idioms like *call the shots*. The schema encodes your knowledge of VPs, and that makes it easier to learn new VP idioms.

In what form do you store all these things? In a traditional view of the lexicon, anything that’s regular or redundant is squeezed out and accounted for by rules, so that all that’s left to store is a distillation of exceptionality. Anna Maria Di Sciullo and Edwin Williams call the lexicon “a collection of the lawless.” But this doesn’t really work. Think about *call the shots*. Its VP structure is redundant, because it’s regular. The words are redundant, aside from their meaning. And even the plural marker is redundant – it’s the good old regular plural. So on this view, what’s left to store is an eviscerated husk attached to an exceptional meaning. The situation is even worse with the clichés, which mean exactly what they should. In principle you shouldn’t have to store anything, but in fact it turns out that you do.

In our work on morphology, Jenny Audring and I conclude that all these things are stored in memory in full, redundantly. However, they are partially predictable because they have SAME-EXCEPT links to other items with which they share structure. This includes relation to schemata. For instance, *call the shots* is linked to the VP schema, and this link encodes the fact that the idiom has the same syntactic structure as the schema.

If we look at this as a whole, the notion of *word* isn’t essential to lexical storage. Words are just a particular size of linguistic unit. We store all kinds of things that are larger than words, like idioms and clichés, and we store all kinds of things that are smaller than words, like the plural *s* and the derivational ending *tion*. We also store *schemata* for units larger than words, such as the VP schema, as well as schemata for units smaller than words, such as the phonotactic templates that define possible syllables. Above all, there’s just a huge amount of stored lexical material, all in the same format as words, namely as pieces of phonology, syntax, and semantics.

Let’s think how this conception of the lexicon might apply to music. After typical exposure to music in our culture, you probably know a couple thousand popular songs, folk songs, children’s songs, advertising jingles, and – at least among present company – hundreds of pieces of music varying in length from five minutes to a couple of hours or more. What I mean by knowing them is if you turned on the radio in the middle of one of them, you could identify it within a few seconds or at least recognize it as familiar. (Carol Krumhansl has a paper documenting this ability experimentally, for popular music.) So all this music has to be stored in some form or another.
But, following the example of the linguistic lexicon, we can also think of storing more abstract knowledge in the musical lexicon as well: the melodic clichés, the forms of cadences, characteristic accompaniment patterns such as the Alberti bass, and so on. Somewhat more abstract and somewhat larger are Robert Gjerdingen’s galant schemas, as analyzed in Fred’s TPS.

And we can store even larger and more abstract patterns too, such as the 12-bar blues, the 32-bar popular song form, and the normative prolongational structure of GTTM and TPS – these are still chunks of musical structure. We can also store very small-scale materials as well: basics of the tonal system such as the relative stability of a particular scale step in a particular scale, and the basic principles of well-formedness for each of the GTTM levels and their interactions – again, chunks of musical structure, just of smaller size. This parallels the situation in the linguistic lexicon – but without a counterpart of words.

What does this say about how tunes themselves are stored? You might think that, like the traditional lexicon in linguistics, you store only the non-predictable parts. But this won’t work. What would it be like to take a tune, remove from it all traces of meter, which is redundant, and remove all traces of tonality, which is redundant, and remove all traces of melodic repetition, which is also redundant? You’d be left with incomprehensible junk. A better story is that you store a tune in full (at least to the extent that you remember it), and you also establish same-except links to all the schemas that help support its structure. In addition, if you’ve picked up motivic parallelisms within the piece, you can store the internal same-except links that register the parallelism. So in the end, your knowledge of a piece ends up as this very rich network, not just a sequence of notes.

In parallel with linguistic memory, a musical representation can be associated with something like a degree of strength or resting activation, correlated with how well you know the passage in question. This lexical strength may depend on its frequency in your experience but probably with other factors as well. If you can recognize a piece but you can’t sing it or play it on demand, presumably its strength is less than if you know it “by memory” and can reproduce it. Somewhere in between might be when you can sing or play the piece from the score but you can’t do it by heart.

Let me push the parallel to language processing a little further. A big question in psycholinguistics is the balance between storage and computation. For example, when you recognize the word construction, do you access the whole word construction, stored in memory? Or do you take it apart into construct plus the stored schema for the suffix -tion, and recognize it as the sum of these parts? The answer I subscribe to is that the brain tries both ways at once. If they both come up with the same answer, that’s even better: thanks to the redundancy, the brain’s response is more robust and reliable.

How would this play out in music? The idea is that when you’re hearing music, you activate all the representations in memory that match what you’re hearing. The crucial thing is that this includes all the relevant schemas! If it’s a piece you don’t know, the schemas have to do all the work of building the piece’s structure. If it’s a piece you do know, even just to some extent, then both the schemas and your knowledge of this particular piece jump in and collaborate in building the structure. That should make your hearing richer.
This view of activation leads to the last point I want to discuss, the notion of musical expectation or prediction. This has been a staple of music theory, and it’s been central to the work of several of today’s speakers (Carol, Lisa, and Davy), among many others. More recently, the notion of predictive processing has taken root in psycholinguistics – it’s interesting that music theory got there first!

In psycholinguistics, the idea is that someone hearing spoken language is making probabilistic predictions about what is coming up – both concrete predictions like what word is coming up, and more abstract predictions like what syntactic structure is coming up.

Now, let’s put this in the context of the perspective on memory I’ve been describing. There are two questions that are obvious once you think to ask them: first, what is the form of a prediction? and second, where do the predictions come from? What’s generating them?

The answer to the first question, both for language and for music, is that the form of a prediction has to be a piece of structure. In music, as in language, this structure may be very concrete, say the next note or chord, as in Carol Krumhansl’s classic experiments. Or it could be more abstract, say a prediction that a full cadence is coming four beats from now, or even more abstract, say, that I’ve heard a minuet and a trio has begun, and so I expect or predict a reprise of the minuet – a large-scale narrative structure.

And where do the predictions come from? Well, what you’ve heard so far activates schemas that build partial candidate structures. The brain doesn’t wait till the end of a schema to activate it. Just hearing the beginning – the left half – is enough to engage it. But once it’s activated, the rest of it, the right half, is active as well, and so the schema is expecting or predicting that this portion that hasn’t been heard yet is coming up: This is how it’s going to go on. Since there are schemas at all sizes of structure and all degrees of abstraction, predictions will likewise vary in their size and their specificity.

In language, it’s usually the case that multiple candidate words and schemas are active and in competition with one another. I can imagine a similar way of thinking about probabilistic predictions in music: multiple schemas with different predictions are in competition, and the probability of each prediction corresponds to the strength of activation for the relevant schemas.

What if it’s a piece that you know? Then your memory for the piece itself – and its structure – best matches what’s heard so far, and it dominates the candidate predictions. In fact, it is so strong that it can even trigger auditory images of what’s coming next, before it comes – I have this experience all the time, and I’m guessing you do too. Still, the schemas are still active, in addition to the piece itself, so a strong schema like the perfect cadence can still make predictions that are thwarted by a deceptive cadence or some other tactic of avoidance.

Well, I’ve certainly not done enough homework to go beyond this very preliminary sketch. But I hope I’ve convinced you that in all these phenomena there are interesting parallels between the theories of language and of musical cognition. They’re not the usual parallels that have been drawn, largely because they come from a different conception of how language works,
the Parallel Architecture framework. But I think these parallels work much better than the standard ones. They’re all challenges for future research.

Now that Fred and I are both retired, who knows, maybe we’ll start thinking about some of these things again?

Anyway, congratulations to Fred on a great career – so far! – and all my best wishes for the next phase of life.

Thank you!