Musical Meaning within Super Semantics[*]

Philippe Schlenker[**]

January 17, 2020

Music consultant: Arthur Bonetto

Abstract. As part of a recent attempt to extend the methods of formal semantics beyond language ('Super Semantics'), it has been claimed that music has an abstract truth-conditional semantics, albeit one that has more in common with iconic semantics than with standard compositional semantics (Schlenker 2017, 2019a, b). After summarizing this approach and addressing a common objection (here due to Leonard Bernstein), we argue that music semantics should be enriched in two directions by incorporating insights of other areas of Super Semantics. First, it has been claimed by Abusch 2013 that visual narratives make use of discourse referents akin to those we find in language. We argue that a similar conclusion extends to music, and highlight it by investigating ways in which orchestration and dance make cross-referential dependencies more explicit. Second, it has been claimed that co-speech gestures trigger characteristic conditionalized presuppositions, called 'cosuppositions', and that their semantic status derives from their parasitic character relative to words (Schlenker 2018a,b). We argue that the same conclusion extends to some instances of film and cartoon music: it may trigger cosuppositions that can be revealed by embedding film excerpts or gifs in sentences so as to test presupposition projection. We further argue that under special discourse conditions (pertaining to certain Questions under Discussion), pro-speech gestures and pro-speech music alike can trigger cosuppositions as well. These results suggest that new insights can be gained not just by extending the methods of semantics to new objects, but also by drawing new connections among them.

Keywords: music, music semantics, musicology, anaphora, cosuppositions, picture semantics, visual narratives, co-speech gestures, co-film music, co-gif music, pro-speech music, co-speech gestures, pro-speech gestures

1   Introduction[1]

There is by now a well-established tradition of study of musical syntax with formal means (e.g. Lehrdahl and Jackendoff 1983, Lerdahl 2001, Pesetsky and Katz 2009, and Rohrmeier 2011 for classical music, Granroth-Wilding and Steedman 2014 for jazz). The formal study of musical meaning is a more recent and more controversial endeavor, in part because its very object is in doubt: does music genuinely have meaning? By meaning, we have in mind *a rule-governed way in which music can provide information (i.e. license inferences) about some music-external reality*, no matter how abstract.[2] Building on numerous earlier insights, both introspective and experimental, it was recently proposed that a music semantics can be developed (Schlenker 2017, 2019a, b, Migotti 2019, Migotti and Zaradzki 2019). This is part of a more general attempt to apply the general methods of formal semantics beyond human language (called 'Super Semantics' or sometimes 'Formal Semiotics', see for instance Greenberg 2013, Abusch 2013, 2015, Schlenker 2019b). First, there are systematic ways in which music triggers inferences about a music-external reality: music semantics has an object. Second, these inferences can be captured by a theory of musical truth, based on the idea (due in part to Bregman 1994) that *the informational content derived from a musical piece is given by the inferences one can draw about its virtual sources*[3]. Musical inferences are of two types: some are lifted from normal auditory cognition, as when a diminuendo sound (decreasing in loudness) is taken to signal that the virtual source is losing energy or moving away from the perspectival point. Other inferences are triggered by specifically musical properties, as when a dissonance is taken to signal that the virtual source is a physically or emotionally unstable position. By positing very simple rules of preservation of various musical properties (such as loudness or harmonic stability) in a space of denotations, a 'proof of concept' was proposed for a truth-conditional music semantics.

This piece has three goals. We start by summarizing for a semantics audience the main claims of music semantics (Section 2), and we address a common objection according to which music has no meaning besides the emotions it triggers in the listener (Section 3). We address this objection on the example of Leonard Bernstein's famous claim that even program music (i.e. music composed to evoke concrete scenes) doesn't have anything like the meaning it purports to have: we revisit one of his own examples (pertaining to Strauss's Don Quixote) and show that in fact it supports the view that music has a truth-conditional semantics, although a far more abstract one than is postulated in program music. This example will also help establish the fruitfulness of the method of minimal pairs, whereby musical snippets can be 'recomposed' to assess the reality and source of various inferences.

We then argue that new insights into musical meaning can be gained by drawing connections with other parts of Super Semantics. First, it has been claimed by Abusch 2013 that visual narratives make use of discourse referents akin to those we find in language. We argue that a similar conclusion extends to music, and highlight it by investigating ways in which orchestration and dance make patterns of cross-reference more explicit (Section 4). Second, it has argued that co-speech gestures trigger characteristic conditionalized presuppositions, called 'cosuppositions', and that their semantic status derives from their parasitic character relative to words (Schlenker 2018a, b). We argue that the same conclusion extends to some instances of film and cartoon music, which triggers cosuppositions that can be revealed by inserting snippets of films or gifs (i.e. a very brief silent cartoon[4]) in a sentence in order to test presupposition projection (Section 5). Finally, we argue that a further finding of gestural research can arguably be replicated in music: under specific pragmatic conditions, pro-speech gestures (which fully replace words rather than accompanying them) can also trigger cosuppositions; we argue that the same conclusion might apply to some examples of pro-speech music.

Taken together, these results suggest that new insights can be gained not just by extending the methods of semantics to new objects, but also by drawing new connections among them.

---

[1] Audiovisual examples have been included by way of URLs (some are borrowed from Schlenker 2017, 2019a,b). We recommend that the reader consult them for the text will be hard to follow in their absence.

[2] This notion of semantics corresponds to what Koelsch 2012 calls 'extra-musical meaning'.

[3] The term 'virtual source' is due to Bregman, e.g. Bregman 1994. See also Nudds 2007 for an analysis of auditory cognition in terms of source perception.

[4] *Gif* just stands for *Graphic Interchange Format*, but the term has come to be used to refer to brief moving images.

2   Music semantics: a source-based analysis

We start by summarizing the main ideas of the formal music semantics proposed in Schlenker 2017, 2019a,b, and briefly mention some objections and extensions proposed by Migotti 2019 and Migotti and Zaradzki 2019. For the sake of clarity and brevity, we restrict attention to a "bare-bones" music semantics. (More sophisticated notions pertaining to the interface between musical syntax and semantics, or semantics and pragmatics, are discussed at varying levels of detail in Schlenker 2017, 2019a.)

2.1   Main ideas

Recent music semantics treats music as a kind of abstract auditory animation. Just as would be the case with auditory or visual perception, the subject seeks to find information about the causal sources of her percepts. In auditory perception, certain sounds reach the human ear and, depending on their properties, give rise to information about the surrounding objects and events: one may for instance hear a car engine moving away as its loudness decreases. The same general idea applies to music semantics: the listener seeks to draw inferences about certain sources. But music semantics is special in several respects (see Schlenker 2019a for a more detailed discussion).

First, the sources are more abstract that actual sound sources (the cellist, the orchestra, or even the conductor): in Bregman's terminology, they are "virtual sources". Still, for him, "the virtual source in music plays the same perceptual role as our perception of a real source does in natural environments". As a result, "transformations in loudness, timbre, and other acoustic properties may allow the listener to conclude that the maker of a sound is drawing nearer, becoming weaker or more aggressive, or changing in other ways" (Bregman 1994).

Second, inferential rules are of two kinds. Some are, in accordance with Bregman's idea, lifted from normal auditory cognition. But others are more specifically musical in nature; this particularly applies to harmonic notions. In Western classical music and in jazz, a key notion is that of a tonal pitch space, with parts that are more stable than others, areas that correspond to 'keys', and non-trivial relations of distance among notes or chords. Across cultures, the special case of a 'tonic center' (a note or chord of greatest stability) appears to play a role as well (Mehr et al. 2019). Lerdahl 2001, 2019 hints at an analysis of musical meaning in terms of a "journey through tonal pitch space", while Ganroth-Wilding and Steedman 2014 provide an explicit semantics for jazz sequences in terms of motion in tonal pitch space.

These inferential rules can be illustrated as follows:

To see a very simple example, both kinds of inferences can be used to signal the end of a piece. One common way to signal the end is to gradually decrease the loudness and/or the speed. While this device could be taken to be conventional, it is plausible that it is in fact derived from normal auditory cognition: a source that produces softer and softer sounds, and/or produces them more and more slowly, may be losing energy.[5] But on the tonal side, it is also standard to mark the end of a piece by a sequence of chords that gradually reach maximal repose, ending on a tonic. Plausibly, an inference is drawn to the effect that a virtual source that manifests itself by a tonic is in the most stable physical position, with no tendency to move any further. Thus these two types of inference combined conspire to signal the end of a piece. (Schlenker 2019a)

A list of 9 examples of inferential effects appears in simple form in Schlenker 2019b (Appendix II); further examples are discussed in Schlenker 2017, 2019a.[6] We provides examples in (2) to show that

---

[5] While the notion of 'energy' should be further explicated, we can rely at this point on an intuitive notion of folk psychology, according to which objects are taken to have different levels of energy depending on their movements and more generally on their behavior.

[6] Bedoya 2019 tests and uncovers further inferential means within the area of musical emotions In a nutshell, he starts from properties of the human voice that are indicative of certain emotions. For instance, a person speaking that starts to smile will produce slightly different sounds, in such a way that one can 'hear' the person smile (formants will be shifted upwards). Bedoya then runs algorithms that perform the same modifications on musical snippets, thus artificially producing "smiling violins", for instance. Finally, he tests the effect of the modification on the emotions conveyed by the music: smiling violins were thus taken to express more positive emotions than

musical inferences have real substance, but also to highlight that their general form is that in (1), which pertains to the preservation of some relations among musical events in the space of denoted (i.e. world) events. This will matter shortly.

(1)   If musical events M_1 and M_2 stand in relation R (M_1 R M_2), their respective denotations e_1 and e_2 stand in relation R* (e_1 R* e_2). This will matter shortly.

(2)   Examples of inferential effects (Schlenker 2019b Appendix II, with links to examples)[7]
       a. Lower pitch may indicate that a virtual source (i) is larger, or (ii) is less excited/energetic.
       b. Lower loudness may indicate that the source is (i) less energetic, or (ii) further away.
       c. Lower speed may indicate that the source is slower.
       d. Silence may indicate that an event is interrupted.
       e. Lesser harmonic stability may indicate that the source is in a less stable (i) physical or (ii) emotional position.
       f. A change key may indicate that the source is moving to a new environment.

Third, and relatedly, within this source-based semantics music can trigger inferences about objects that are not sound-producing. This abstractness is essential in view of the fact that music can help evoke silent scenes (an example among many others can be found in Saint Saëns's piece The Aquarium in his *Carnival of the Animals*: the creatures evoked are not or barely sound-producing).[8] This result is achieved because the inferential mechanisms at play allow sound properties to produce information about properties of virtual sources that are not implicated in sound production. For instance, decreasing loudness may be interpreted in terms of a source losing energy or moving away, and both events may characterize entities that are silent. This point will become clearer when we illustrate the workings of the proposed semantics, to which we now turn.

2.2   *Schematic illustration*[9]

As a proof of concept, Schlenker 2017, 2019a discusses a very simple theory of musical truth in which interpretation explicitly proceeds by requiring that some relations among musical events be preserved in the space of the denoted (world) events. In other words, the requirements are precisely of the form defined in (1) and illustrated in (2). For greatest simplicity, we consider three musical events, as in (3), which each have only two properties: loudness is interpreted in accordance with (2)b (= lower loudness corresponds to lower energy or greater distance of the source), while harmonic stability is interpreted in accordance with (2)e(i) (= lesser harmonic stability corresponds to a lesser physical stability of the source). These two properties were chosen to show that the proposed semantics makes it possible to aggregate very diverse inferences, with some lifted from normal auditory cognition, while others are more intrinsically semantic in nature.

The three musical events in (3) are each defined by a pair of two properties. As in music theory, *I* refers to a tonic chord, which is harmonically maximally stable (this is the chord CEG in the key of C), while *V* refers to a dominant chord (= GBD in the key of C), which is a bit less stable.

(3)   M =   <<I, 70db>, <V, 75db>, <I, 80db>>

Since I is more harmonically stable than V, the first and third denoted events (corresponding to the initial and final tonic chord I) should be more stable than the second one. The three-chord sequence features a crescendo, with loudness going from 70db, to 75db, to 80db; correspondingly, the three events should either correspond to a source that gains energy, or one that approaches the perspectival point.

The basic intuition of the framework is that musical voices (i.e. melodically coherent individual parts of a piece) are associated with virtual sources that are objects, and participate in certain events.

---

standard violins. Manipulations studied included pitch (tuning up vs. tuning down), formants (smiling vs. "unsmiling" music), vibrato (frequency modulations around the base frequency) and "roughness".

[7] As noted in Schlenker 2019b, this list was originally prepared for an interview incorporated in Keats 2018.

[8] The beginning of Richard Strauss's Thus spoke Zarathustra, discussed in Schlenker 2017, 2019a, was for its part intended to evoke a sunrise.

[9] This summary follows closely that of Schlenker 2019b.

Therefore a voice involving n musical events was taken to denote a pair of an event and of n real world events, as is stated in (4).

(4)   Let M be a voice, with M = $\langle M_1, \ldots, M_n \rangle$. A possible denotation for M is a pair $\langle O, \langle e_1, \ldots, e_n \rangle \rangle$ of an object and a series of n events, with the requirement that O be a participant in each of $e_1, \ldots, e_n$.

Starting from the musical sequence in (3) and the specification of possible denotations in (4), it is then posited that the musical piece M = $\langle M_1, \ldots, M_n \rangle$ is true of the pair of an object O and events it takes part in, $\langle O, \langle e_1, \ldots, e_n \rangle \rangle$, just in case $\langle O, \langle e_1, \ldots, e_n \rangle \rangle$ is a possible denotation for M, and in addition the mapping from $\langle M_1, \ldots, M_n \rangle$ to $\langle e_1, \ldots, e_n \rangle$ preserves certain requirements, listed in (5). Informally: the denoted events should preserve the temporal ordering of the musical events, as well as the loudness and stability ordering among them.

(5)   **Defining 'true of' in music**
Let M = $\langle M_1, \ldots, M_n \rangle$ be a voice, and let $\langle O, \langle e_1, \ldots, e_n \rangle \rangle$ be a possible denotation for M. **M is true of $\langle O, \langle e_1, \ldots, e_n \rangle \rangle$** if it obeys the following requirements.

a. Time
The temporal ordering of $\langle M_1, \ldots, M_n \rangle$ should be preserved, i.e. we should have $e_1 < \ldots < e_n$, where $<$ is ordering in time.

b. Loudness
If M_i is less loud than M_k, then either:
(i) O has less energy in e_i than in e_k; or
(ii) O is further from the perceiver in e_i than in e_k.

c. Harmonic stability
If M_i is less harmonically stable than M_k, then O is in a less stable position in e_i than it is in e_k.

To illustrate, Schlenker 2017, 2019a, b considers different kinds of events made of three subevents: a sunrise and a sunset (with the sun as virtual source gaining or losing luminosity), a boat approaching or departing (with the boat as moving source moving closer or further away). The rules as stated

make it possible to take the sequence M in (3) to be true of a sunrise involving three subevents: minimal luminosity, rising luminosity, maximal luminosity. The apparent energy of the source rises, as mandated by the Loudness condition; and the first and third subevents are more stable than the second one, as mandated by the Harmonic stability condition (this is on the assumption that events of 'minimal luminosity' and 'maximal luminosity' involve little or no change, whereas 'rising luminosity' involves a faster change). By contrast, a sunset would fail the Loudness condition, as the apparent level of energy of the source does not rise. Similarly, interpreting the Loudness condition in terms of proximity rather than in terms of level of energy, the same sequence could be satisfied by a boat approaching, with three subevents: maximal distance, movement towards the source, minimal distance (here too, with the assumption that the first and last event are more stable that the second). By contrast, a boat *departing* could not satisfy the Loudness condition.

It is worth noting that these preservation conditions are abstract enough that they can be satisfied by real world events that are not sound-producing (such a sunset or a boat approaching), and may be very diverse; this is the sense in which musical meaning is in general *very* abstract.

2.3   *Extensions*

Schlenker 2017, 2019a discusses at various levels of detail theoretical issues pertaining to the interface between (ii) syntax and music semantics, and (ii) semantics and pragmatics. Important objections and refinements were also proposed by Migotti 2019 and Migotti and Zaradzki 2019.

Migotti 2019 correctly noticed that interpretation by way of preservation rules akin to those in (5) is too permissive. Consider again (3), and assume that the three chords are played at regular intervals of 1 second. Preservation of ordering in time would allow the first two denotations to be separated by one minute, while the second and third are separated by one day – which has no plausibility at all. Time preservation is probably far stricter: if we view music as an abstract auditory animation, it is likely that time is often interpreted without change (i.e. a 1 second interval between the notes is interpreted as a 1

second interval between the denoted events), or possibly with a multiplicative parameter that remains relatively constant throughout a passage.

But the point is more general. In a detailed study of musical excerpts that strongly evoke a person walking, Migotti and Zaradzki 2019 note that an alternation between a stable and a slightly less stable chord is maximally effective; but it is essential that the stable chord be *absolutely* stable and not just more stable than the less stable chord. This is unexpected if all that matters is the preservation of certain orderings. The general objection is no doubt right: stronger preservation principles ought to be explored. Migotti 2019 considers preservation *modulo* a multiplicative parameter, and he sketches for loudness a more ambitious analysis in which the details of the inferential rule are derived from the physics of sound, in the sense that the inferences that one draws on the level of energy of the source are determined by those that are in fact physically licensed.[10]

We will now disregard these issues to focus on an objection due to the great composer and conductor Leonard Bernstein. This will have two benefits: to address head-on a fundamental problem, and to illustrate the main ideas of a music semantics on a very concrete example.

3   An objection to music semantics: Bernstein's challenge

3.1   'No semantics'

There is a long tradition of scholars denying that music conveys information about the extra-musical world. Different views converge on this conclusion. One, due for instance to Hanslick (1891), is that music just has no "subject matter".[11] Another is that music only has an internal semantics, in the sense that it triggers certain inferences and expectations about its own form, which in turn may trigger certain emotions. To cite but two examples of an internal semantics, Meyer 1956 writes that "one musical event (...) has meaning because it points to and makes us expect another musical event" (Meyer 1956, chapter I); this gives rise to expectations and emotions that constitute what Meyer calls "embodied meaning". Huron 2006 argues that various emotions of a musical or extra-musical nature derive from general properties of expectation, or in other words of our attempts to anticipate what will come next, in music or elsewhere.[12] Yet another common view is that the meaning of music entirely lies in the emotions it evokes in the listener. We discuss it in greater detail in the concrete version that was articulated with panache by Leonard Bernstein.

3.2   Bernstein's objection[13]

In his celebrated 'Young People's Concerts', Leonard Bernstein devoted an entire program to 'What is Musical Meaning?' (1958; see also Bernstein 2005), and he argued that the true meaning of music is

---

[10] Migotti and Zaradzki 2019 also raise a more fundamental issue. They argue in a study of walk-denoting excerpts that a 2-chord sequence enriched with a third, less salient note may evoke a walk even though no subevent seems to correspond to that third note. If this is indeed the case, one possibility they sketch is that certain notes play a role roughly similar to that of modifiers in language: they modify the interpretation of the notes they accompany but do not represent an event on their own. An alternative (which may or may not turn out to be a notational variant) would be to take the granularity of interpretation to be somewhat variable, possibly with cases in which a *group* of notes is taken to denote an event (as discussed speculatively in Schlenker 2019, Appendix IV).

[11] More specifically, Hanslick 1891 writes: "while sound in speech is but a sign, that is, a means for the purpose of expressing something which is quite distinct from its medium; sound in music is the end, that is, the ultimate and absolute object in view." (Hanslick 1891 p. 94) Later in the same piece, he writes: Music has (…) no subject in the sense that the subject to be treated is something extraneous to the musical notes"(Hanslick 1891 p. 162)
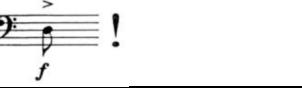
[12] For Huron, "the emotions evoked by expectation involve five functionally distinct physiological systems: imagination, tension, prediction, reaction, and appraisal" (p. 7), and he tries to derive musical emotions from the interaction of these systems with musical anticipations (the resulting theory is called 'ITPRA', which is the acronym of the five physiological systems).

[13] Thanks to Paul Egré (p.c.) for calling our attention to the relevance of Bernstein's discussion for music semantics.

"the way it makes you feel when you hear it".[14] His argument was that even purportedly referential ("program") music doesn't convey information about the world. As a case study, he discussed Variation II of Richard Strauss's Don Quixote, and showed that one can tell the *wrong* story and still have something that fits the music just as well as the 'real' story. To get his point across, Bernstein had his orchestra play the Strauss piece to illustrate a story he told about Superman. Then the orchestra played the music again, but now to illustrate an episode of Don Quixote, in accordance with Strauss's intentions. Bernstein's point was that the Superman interpretation worked just as well as the original interpretation.

As briefly noted in Schlenker 2019b (Appendix II), Bernstein's point can in fact help bring out what music semantics is about. Bernstein is clearly right about two basic facts. First, a naive subject who listens to Variation II would be hard pressed to guess almost any of the story – contrary to someone who saw a visual depiction of the same story. Second, the music can indeed be made to fit a different story, such as that of Superman as told by Bernstein. But most strikingly, Bernstein's Superman interpretation is almost entirely isomorphic to the Don Quixote interpretation. We reproduce in (6) the correspondence. Bernstein's point doesn't show that music doesn't have a semantics; rather, it beautifully illustrates the fact that music has an *abstract* semantics.

(6) **Simplified structure of Bernstein's Don Quixote and Superman interpretations of Strauss's Variation II of *Don Quixote*** (Kriegerisch. "Der siegreiche Kampf gegen das Heer des großen Kaisers Alifanfaron" ("The victorious struggle against the army of the great emperor Alifanfaron") [actually a flock of sheep]) Entire discussion: https://youtu.be/XFZ7wORtj2A (links from Schlenker 2019b)

Don Quixote interpretation	Superman interpretation	Salient musical passage
Context: Don Quixote is a foolish old man who has read too many books about knighthood and decides he is a marvelous knight himself. Sancho Panza is his devoted servant. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=5m17s	**Context:** An innocent man can't sleep in a prison where he was put unjustly. He spends his night playing the kazoo while other prisoners snore. But his friend Superman is coming to rescue him. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=28s	
Don Quixote departs on his horse to conquer the world. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=5m36s	Superman comes charging along through the alley on his motorcycle. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=1m8s	
We hear Sancho chuckling to himself[15]. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=5m45s	Superman whistles his secret whistle (in the woodwinds) so the prisoner will know he's coming. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=1m20s	
They see a flock of sheep in the field going *baa-baa*. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=6m3s	Superman hears all the prisoners snoring away peacefully in the dead silence of night. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=1m28s	
A shepherd is playing on his pipe. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=6m16s	Over this snoring, Superman hears his imprisoned friend playing his kazoo over the snoring, which gets louder as he gets nearer. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=1m50s	
Don Quixote charges at the sheep, taking them to be an army. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=6m27s	Superman charges into the prison yard and bops the guard over the head, done in the orchestra with a loud bang in the percussion. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=2m14s	loud bang in the percussion:
The sheep run off in all directions baaing wildly. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=6m40s	The kazoo stops playing, and with all the snoring still going on, Superman grabs his friend and carries him away on his motorcycle. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=2m22s	

---

[14] Bernstein revisited this topic with different views in his Harvard Lectures (Bernstein 1976): "music has intrinsic meanings of its own, which are not to be confused with specific feelings or moods, and certainly not with pictorial impressions or stories. These intrinsic musical meanings are generated by a constant stream of metaphors, all of which are forms of poetic transformations." We focus on the Young People's Concerts for their rich empirical content (and clarity) rather than for the positive theory Bernstein develops in them.

[15] The text has "chuckling to himself", Bernstein's live performance has: "laughing at Don Quixote" (there are several small differences between the live and the printed version).

	The snoring gets farther and farther away, until we don't hear it any more. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=2m37s	
Don Quixote is convinced he has done a truly knightly deed, and is he proud! https://www.youtube.com/watch?v=XFZ7wORtj2A&t=6m45s	Our hero at last reaches freedom! https://www.youtube.com/watch?v=XFZ7wORtj2A&t=2m50s	

As illustrated in (7), the correspondence is almost complete. Don Quixote departing, charging the sheep, and triumphing corresponds to Superman leaving on his motorcycle, charging into the prison and triumphing ((7)a, d, g). The sheep going *baa-baa* corresponds to the prisoners snoring ((7)c, f). And the shepherd playing on his pipe gets reinterpreted in terms of the prisoner playing on his kazoo ((7)d). The only structural difference is that Sancho Panza and Don Quixote are merged in the Superman interpretation, with the result that Sancho Panza chuckling is reinterpreted as Superman whistling his secret tune ((7)b). Structurally, this is virtually the only difference between the two stories.[16]

(7)   **Correspondence in terms of sources between Bernstein's Don Quixote and Superman interpretations**

Don Quixote interpretation	
a. **Don Quixote** departing on his horse	**Superman** charging along on his motorcycle
b. **Sancho chuckling**	**Superman whistles**
c. **Sheep** going *baa-baa*	**Prisoners snoring** away peacefully
d. A **shepherd** is playing on his pipe	**Imprisoned friend** is playing his kazoo.
e. **Don Quixote** charges at the sheep	**Superman** charges into the prison yard
f. The **sheep** run off baaing wildly (and become more distant)	With the **snoring** still going on, Superman carries his friend away.
g. **Don Quixote** is convinced he has done a truly knightly deed, and is he proud!	**Superman** (with his friend) at last reaches freedom!
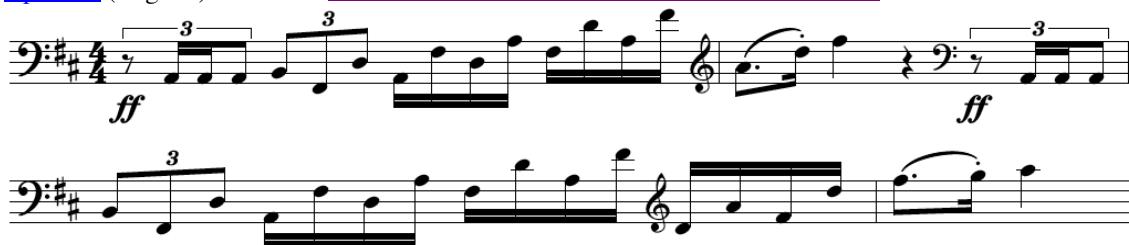
3.3   *Music semantics in action: minimal pairs*

The point made above (following Schlenker 2019b, Appendix II) is too weak, however: it could be that Bernstein just didn't pick the optimal story to show that music lacks a semantics. But as we will now see, salient musical effects of Strauss's Variation II can be shown to have genuine semantic implications, ones that are abstract yet greatly constrain the space of possible denotations; this, in turn, suggests that not anything goes when one seeks to tell the 'wrong' story to fit a musical piece: salient musical effects that give rise to inferences will have to be properly interpreted by the story, and hence different acceptable stories will likely have a lot of structural properties in common.

3.3.1   *Rising frequency*

We start with a use rising frequency to evoke a rise in energy, as in (2)a. This is used to evoke Don Quixote's or Superman's triumphant departure in Bernstein's stories, as shown in (8).

(8)   Upwards (original)   https://www.dropbox.com/s/dn9cwtc0zf7ji9m/Strauss-Don%20Quixote%20cello%20original.aiff?dl=0

---

[16] As noted in Schlenker 2019b about (7)f, "the musical chaos corresponding to the sheep's baaing wildly is not easily interpreted in the Superman story (why would the prisoner's snoring become more chaotic when Superman grabs his friend and liberates him?)".

To test the contribution of the rising frequency to the evocation of a triumphant departure, we create (thanks to Arthur Bonetto's help) a minimal pair that inverses all the melodic motions according to rules of tonal composition.[17] Strikingly, the result is musically acceptable music, but the abstract inferences it triggers are completely different from those of the original: the general impression of a triumphant departure has been destroyed.

(9) Downwards (A. Bonetto) https://www.dropbox.com/s/xffokt0huvkg1fl/Strauss-Don%20Quixote%20cello%20vers%20le%20bas.aiff?dl=0

3.3.2 Dissonances

Next, we turn to dissonances used to evoke a flock of sheep going *baa-baa* in the Don Quixote story, and prisoners snoring in the Superman story. As it happens, this is a case in which the musical dissonances are used to evoke chaotic events that are themselves sound-producing, although it can easily be checked that even the orchestral version doesn't really resemble sheep baaing or prisoners snoring.

(10) **Dissonances evoking chaos** (temporal alignment plays a role too)

They see a **flock of sheep** in the field going *baa-baa*. https://www.youtube.com/watch?v=XFZ7wORtj2A&t=6m3s	Superman hears all **the prisoners** snoring away peacefully . https://www.youtube.com/watch?v=XFZ7wORtj2A&t=1m28s

The dissonances are produced by multiple chords that contain notes that are only one half-tone apart, as shown in the boxed parts of (8). When the music is rewritten so as to minimally remove the dissonances, as in (12), this impression of chaos almost entirely disappears.[18]

(11) Dissonances (original, simplified Midi) https://www.dropbox.com/s/memrvak6ewewosi/Strauss-Don%20Quixote%20moutons%20original%20v2.aiff?dl=0

(12) No dissonances (A. Bonetto) https://www.dropbox.com/s/f76s0acuv3mwu26/Strauss-Don%20Quixote%20moutons%20sans%20dissonances%20v2.aiff?dl=0

---

[17] Technically, the recomposition was effected by taking symmetric intervals relative F# (= the 3rd degree in the relevant key, namely D major). To illustrate, the first note of (8) is A. The mirror-image note relative to F# (going downwards) is D, as in (9). The second degree in (8) is B. Its mirror-image counterpart relative to F is C#, the second degree that appears in (9).

[18] Bonetto kept the same number of notes in each chord, finding the closest chord that was in the key of D major.

One should not conclude that dissonances in music are solely used to evoke dissonant sounds in nature. There are multiple examples in which this is not so. A particularly simple one was mentioned in Schlenker 2019a. In his *Carnival of the Animals*, Saint Saëns uses a radically slowed down version of the French *Can Can* dance to evoke tortoises (see http://bit.ly/2DAbnrN). Later in the piece, dissonances are suggestive of the tortoises tripping, as in (13)a. The effect entirely disappears when the music is rewritten so as to remove the dissonances, as in (13)b.

(13)   A dissonance is used to evoke tortoises tripping in Saint Saëns's *Carnival of the Animals*
a. In the original version, there is a dissonance in the first half of measure 12 because a chord F A C is played with an G# added (as can be heard by focusing only on the violin and piano parts). http://bit.ly/2ECNWNJ

b. The dissonance can be removed by turning the G#'s into A's – and the impression that tortoises disappears (as can be heard by focusing only on the violin and piano parts). http://bit.ly/2CWFVCT

A dissonance used to evoke an equally silent emotional rather than physical imbalance is used in the music of Hitchcock's *Psycho* http://bit.ly/2mAjZGL. The excerpt in (14)a starts with a D F# Bb (augmented fifth) chord, which sounds dissonant – and is preserved over the first half of the second bar. While other choices contribute to the impression of mental imbalance (such as *ostinato* of the basic melodic movement and the rhythm), the semantic effect is considerably reduced when the dissonances are removed, as in (14)b,c.

(14)   Herrmann's Psycho - reduction, re-written in G minor (A. Bonetto; Schlenker 2019a)[19]
a. Original reduction   http://bit.ly/2D2NlEK
b. Same as in a., re-written in G minor without dissonances   http://bit.ly/2EH4iFt
c. Same as b., closer to the original harmony   http://bit.ly/2mtDXmL

3.3.3   Loudness

The passage of Strauss's Variation II featuring the sheep going *baa baa* (in (10)) also makes use of a crescendo to indicate that a virtual source is approaching, in accordance with (2)b (it doesn't matter for our purposes whether this is because the sheep or the perceiver are moving: movement is relative). In Bernstein's Superman version, the sheep baaing are replaced with prisoners snoring, but the movement is the same.

---

[19] In greater detail, the transformations were as follows:
(i) From (14)a to (14)b: **Bar 1**: F# > G **Bar 2**: F# > G ; B > Bb **Bars 3-4/6-7** : F > G ; Gb > G ; B > Bb **Bar 5:** C > D; B > Bb ; Ab > G ; Eb > D.
(ii) From (14)a to (14)c: same as (i), but the boxed F > G in (i) becomes F > F# instead.

(15) Crescendo evoking the sheep (+ shepherd) approaching

Here too, the semantic effect is easy to diagnose by way of minimal pairs. (16)a displays Bernstein's own interpretation in a celebrated performance in 1943.[20] A simplified piano reduction appears in (16)b, also with a crescendo. The same reduction appears in (16)c, but now with a decrescendo (= decreasing loudness) instead of the crescendo. Instead of an impression that something is approaching, we get the impression that something is moving away.[21]

(16) Minimal modifications (A. Bonetto)
    a. Dissonances < (Bernstein, 1943)   https://www.dropbox.com/s/i61mqotb72kv60i/Strauss-Don%20Quixote-Variation%20II-Berstein-Sheep%20approaching-SHORTER.mp4?dl=0
    b. All <, as in the score (simplified Midi)   https://www.dropbox.com/s/memrvak6ewewosi/Strauss-Don%20Quixote%20moutons%20original%20v2.aiff?dl=0
    c. All > (simplified Midi)   https://www.dropbox.com/s/gyrpkbjts0f2f9p/Strauss-Don%20Quixote%20tout%20en%20dim%20v2.aiff?dl=0

This particular excerpt is remarkable in making a clear use of a crescendo to represent a source approaching, but as discussed in earlier work (e.g. Schlenker 2017, 2019a) and in (2)c, there are multiple cases in which loudness modifications provide information about changes in the state of energy or excitement of a virtual source rather than about its distance from the perceiver.

3.3.4 Cadence

The end of Strauss's Variation II features a cadence evoking a triumphant conclusion. In classical music theory, a cadence is a sequence of chords V - I (dominant - tonic), ending on the most stable tonic chord from the somewhat less stable dominant chord. As discussed at the beginning of this piece (following Schlenker 2017, 2019a), several means are often combined to announce the end of a piece: not just a gradual transition to the most harmonically stable position, but sometimes also a decrease in speed, loudness and even frequency. But what the present analysis leads one to expect is that the end of a piece could have different semantic implications depending on how it is realized. Thus Schlenker 2019a argued that by "considering the interaction between speed and loudness, we can begin to predict how an ending will be interpreted":

---

[20] This was the performance that launched Bernstein's career. As Shawn 2014 writes, "guest conductor Bruno Walter had come down with influenza" and Bernstein had to replace him in a program that included Strauss's Don Quixote. "He had never rehearsed these works with the orchestra, and there wouldn't be time for a minute with them before the performance. Fortunately, he had been fascinated by the complex Strauss score and had painstakingly studied its intricacies and how they mirrored events in the Cervantes novel."

[21] One can also explore more sophisticated minimal pairs in which the melody is played crescendo and the dissonances decrescendo or conversely. The effect is arguably that there are two virtual sources approaching or moving away, as the case may be.

(i) More modifications
[Dissonances <, melody >] (simplified Midi) https://www.dropbox.com/s/klzzuf93savy9zg/Strauss-Don%20Quixote%20moutonscresc%20melodiedim%20v2.aiff?dl=0
[Dissonances >, melody <] (simplified Midi) https://www.dropbox.com/s/s2nhkwue2bo0s8l/Strauss-Don%20Quixote%20moutonsdim%20melodiecresc%20v2.aiff?dl=0

a diminuendo ending can be interpreted as involving a source moving away, or as a source losing energy. In the first case, one would not expect the perceived speed of events to be significantly affected. In the second case, by contrast, both the loudness and the speed should be affected. The effect can be tested by exaggerating the diminuendo at the end of Chopin's Raindrop Prelude in (17); without the ritenuto, the source is easily perceived as moving away.

(17) Last bars of Chopin's Prelude 15 ('Raindrop')
   a. In an exaggerated version of the diminuendo in the normal version, realized with a ritenuto, the source seems to gradually lose energy, becoming slower and softer. http://bit.ly/2CJWHVJ
   b. In a version of a. without ritenuto, the source seems to be moving away, as it gradually becomes softer, without change of speed. http://bit.ly/2qMnRd0

Still, "if we add a crude crescendo instead, and a final accent, the ending sounds more intentional, as if the source gradually gained stamina as it approaches its goal, and signaled its success with a triumphant spike of energy http://bit.ly/2mcPWET."

This last case is closer to what we find at the end of Strauss's Variation II: the ending is realized fortissimo (very loud), and strongly gives the impression of the attainment of a goal. Here we can test the effect played by the final tonic, as in (19), by replacing it with a cluster, a completely dissonant group of haphazard notes, as in (20).

(18) Cadence evoking a triumphant completion

Don Quixote is convinced he has done a truly knightly deed, and is he proud! https://www.youtube.com/watch?v=XFZ7wORtj2A&t=6m45s	Our hero [= **Superman, with his friend**] at last reaches freedom! https://www.youtube.com/watch?v=XFZ7wORtj2A&t=2m50s

(19) Expected chord (I) at the end (original) https://www.dropbox.com/s/jg5u7o8wixexnvv/Strauss-Don%20Quixote%20cadence%20original.aiff?dl=0

(20) Cluster (A. Bonetto)   https://www.dropbox.com/s/vhy5vnlj1bpimii/Strauss-Don%20Quixote%20cadence%20cluster.aiff?dl=0

The effect is unmistakable: one gets the impression that something goes very wrong at the end, possibly Don Quixote falling off his horse or Superman falling off his motorcycle, as the case may be. Any kind of crash would seem to be compatible with the final cluster as well.

3.4 How abstract is musical meaning?

Upon closer inspection, then, Bernstein's example doesn't show that music has no meaning, just that it has an abstract meaning, in the following sense: there are usually lots of very diverse situations that can make a given excerpt true. The striking structural similarity between the Don Quixote and the Superman interpretations of Strauss's variation might initially be thought to be anecdotal. But when we look at the inferences triggered by specific properties of the music, we can check that not anything goes: the rising frequency of the beginning has enthusiastic implications that are radically modified in a minimal modification in which melodic movement is reversed; the dissonances intended to evoke a flock of sheep definitely need not describe sheep, but they do produce an impression of chaos, which is removed if the music is rewritten by eliminating the dissonances; in the same passage, the increasing loudness is naturally interpreted as an object approaching the perceiver (we can't exclude an alternative

interpretation on which it would evoke a rising level of energy of the source, however); and the final, fortissimo cadence is well suited to evoke the attainment of a goal.

This doesn't do justice to the evocative power of the music, but it might be enough to show that there is a vast difference between the view that music has no semantics, and the view that it has an abstract semantics, satisfied by lots of very diverse situations. Bernstein was right to criticize the view that music can paint scenes in the manner suggested by program music. But it doesn't follow that the meaning of music reduces to "the way it makes you feel": all the inferential effects we discussed in this section pertained to what happens in the world, not to the feelings of the listener. (See Schlenker 2019a for a more detailed discussion of the way in which emotions naturally come to play a prominent role in music semantics.)
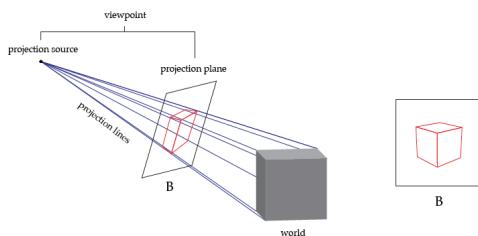
4   Discourse referents in music

Having argued that our music semantics allows for appropriately abstract inferences, illustrated with Strauss's Don Quixote, we turn to an enrichment of music semantics inspired by the semantics of visual narratives proposed by Abuch 2015, building on the pictorial semantics of Greenberg 2013. We start with a comparison between music semantics and pictorial semantics, and argue that a key innovation due to Abusch should be borrowed by music semantics. While Greenberg's theory reduced the meaning of pictures to the set of situations that can be projected onto them, Abusch argued that pictures should be enriched with discourse referents akin to those we use in language (for instance to resolve pronominal reference).[22] As we will see, the argument can be extended to music.

4.1   *Music semantics vs. picture semantics*

Following Schlenker 2019b, we can offer a minimal comparison before a Greenbergian semantics for temporally ordered sequences of pictures, and a music semantics. We start from the notion of pictorial truth in (21), relativizing it to a situation of evaluation. The basic intuition is that a picture is true relative to a viewpoint of those situations that can project onto the picture, as illustrated in (22).
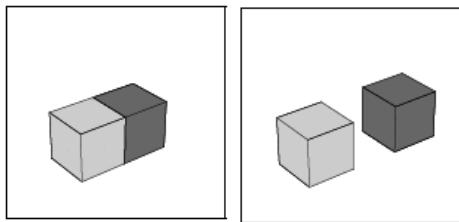
(21) **Truth of a picture (after Greenberg 2018, taking situation rather than worlds to be basic)**
A picture P is true in situation e relative to viewpoint v along the system of projection S iff e projects to P from viewpoint v along S, or in other words: $proj_S(e, v) = P$

(22) **An example of a projection method: perpsective projection (Greenberg 2019)**

One can then extend this notion of pictorial truth to temporally ordered sequences of pictures, as in the case of the 2-picture sequence in (23), from Abusch and Rooth 2017, which represents "a short comic of two cubes moving apart".

---

(23) Picture P1   Picture P2

A very simple notion of truth for n temporally ordered pictures can be given as in (24):

(24) **Picture sequences true of tuples of situations (after Abusch)**
A picture sequence $<P_1, \ldots, P_n>$ is true of situations $<e_1, \ldots, e_n>$ **relative to viewpoint v** along the system of projection S iff
(1) temporally, $e_1 < \ldots < e_n$;
(2) $\text{proj}_S(e_1, v) = P_1$ and $\ldots$ and $\text{proj}_S(e_n, v) = P_n$.

Strikingly, this comes very close to a slightly modified notion of truth for musical excerpts, as in (27). We have added to the definition an explicit auditory point that was implicit in earlier statements, but played an unmistakable role. Specifically, when we stated that lower loudness can be interpreted as a source being further from the perceiver in one event than in the other, as in (5)b(ii), this only made sense because the perceiver played the same kind of role as the viewpoint in (23). For the rest, (27) is similar to our initial definition in (5), except that the role of the object that takes part in the sequence of events has been left implicit. We have kept the vocabulary of events in (27) and of situations in (24).

(27) **Musical sequences true of tuples of events (modified from Schlenker 2017, 2019a,b)**
A musical sequence $<P_1, \ldots, P_n>$ is true of events $<e_1, \ldots, e_n>$ **relative to auditory point v** iff
(1) temporally, $e_1 < \ldots < e_n$;
(2) the Loudness and Harmonic stability conditions are satisfied for the relevant events relative to auditory point v.
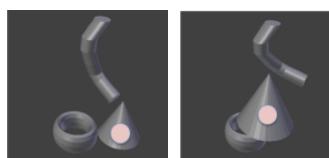
It is clear that part (2) of each definition is entirely specific to the medium: projection rules in pictures have little to do with preservation conditions in music. But in both cases we have animations of sorts (although musical animations are arguably far more continuous than picture sequences, which motivated Migotti's objection to the simplified framework developed here in terms of ordering preservation).

4.2   Adding discourse referents

Abusch 2013, 2015 notices that a definition along the lines of (24) does not do justice to ambiguities that arise in visual narratives, as in the simple example in (28).

(28) **An ambiguity of coreference in pictures** (Abusch 2015)

As Abusch 2015 writes, on a simple picture semantics (28)

is consistent with worlds where a single cone moves in front of a torus. It is also consistent with worlds where the cone of the first picture moves out of view, and another cone moves into view. To infer identity between the cones is to eliminate worlds of the second kind. This is done by adding to the discourse representation a syntactic predication of identity between the two indices, serving the same function as co-indexing in linguistic representations.

Schlenker 2019b argues that this is a genuine ambiguity, not underspecification: embedding the visual narrative under an *if*-clause as in (29) suggests that one naturally obtains a reading on which the cross-reference is resolved.

(29) If what happens next is that , there will be no cone left to close another vase.

Intuitively, "this sentence is true on the salient reading on which it is the same cone that appears in the first and in the second image; it is false on the far-fetched reading on which the first cone comes out of view and a second cone appears." (See Abusch 2013, 2015 and Schlenker 2019b for slightly different ways of introducing variables into pictures.)

We will now argue that a similar move is warranted in music semantics. To motivate it, let us consider again Strauss's Variation II. The cello melody starts as in (30)a, with the triumphant implications discussed above, corresponding to the beginning of the narration mentioned in (7)a. But within a very different background (pertaining to the dissonances evocative of sheep), almost the same melody appears in (30)b, when Don Quixote charges at the sheep, as described in (7)e. Within Strauss's piece, it is clear that the intention is that both melodic lines produced by the cellos evoke not just the same type of triumphant event, but also the very same individual, Don Quixote. Bernstein's Superman story preserves this coference between the two melodies, as can be seen in (7)a,e.

(30) Don Quixote departing vs. Don Quixote charging
      a. **Variation II, beginning, Don Quixote departing** https://youtu.be/A8MChVNcaII

      b. **Variation II, Don Quixote charging** https://youtu.be/V_SKHJFzB5Q

In terms of music semantics, it is clear that nothing as specific as a Don Quixote or a Superman story can be evoked. Still, it would be natural to understand the music as involving the same virtual source in both excerpts. This will be particularly true if the two melodies are played in the same way. Still, this is not something that is absolutely mandated: one could decide to play the two melodies with such different styles (for instance in terms of tempo, dynamics, articulation, maybe even vibrato) that the coreference could be overridden. To put it differently: the performer and the listener are likely to make decisions about the coreference relations that hold among the virtual sources corresponding to different passages. For the performer, these decisions are likely to have important technical and musical repercussions (such as the need to play both melodic excerpts in the same or in different ways).

Since coreference relations are essential but are not entirely determined by the surface of the music, we will posit that musical events (more abstractly: musical voices) can be represented with covert discourse referents (i.e. variables). Accordingly, we propose a revised definition of the semantics along the lines in (31), where each musical event is taken to come with a covert variable that corresponds to the virtual source it represents. Cases of ambiguity will arise when one can plausibly decide to coindex two musical events or not to do so.

(31) **Adding variables to music semantics**
      Let $v_1, \ldots, v_n$ be n variables whose denotations are respectively $D(v_1), \ldots, D(v_n)$. A musical sequence
      $\langle P_1(v_1), \ldots, P_n(v_n) \rangle$ is true of events $\langle e_1, \ldots, e_n \rangle$ relative to auditory point v) iff for each k ($1 \leq k \leq n$), $D(v_k)$ takes part in e_k and
      (1) temporally, $e_1 < \ldots < e_n$;
      (2) the Loudness and Harmonic stability conditions are satisfied (for all the relevant objects and events) relative to auditory point v.

It will be convenient to have at our disposal a refinement of this mechanism in which a musical event can carry a sum of variables. This is motivated in natural language by cases of partial binding, as illustrated in (32), where *they* is bound by the subject quantifier but denotes a group that also includes Mary. One simple way to encode this fact is to allow for complex variables that are sums of simple variables, with the syntax and semantics in (33).

(32)   Mary is very popular in her class. At some point or other, [every boy in her class]$_i$ has asked her$_k$ if they$_{i+k}$ could go on a date. (Büring 2005 p. 197)

(33)   Variables
a. Syntax
(i) for any integers i, x_i is an elementary variable;
(ii) if v and v' are variables, v+v' is a complex variable.
b. Semantics
Let s be an assignment of values to elementary variables, and let $d+d'$ stand for the mereological sum of (possibly plural) individuals d and d'.
(i)for any integer i, if x_i is an elementary variable, its denotation $D(x_i)$ is given by: $D(x_i) = s(x_i)$;
(ii) if $v+v'$ is a complex variable its denotation $D(v+v')$ is given by $D(v+v') = D(v) + D(v')$.

Here an important intuitive principle will matter: different timbres tend to be associated with different objects. This makes good sense in terms of normal auditory cognition: timbre is associated with permanent properties of an object, and thus it is unsurprising than in music semantics different timbres give a strong hint as to the non-identity of different sources. We have stated this principle of interpretation in (34); we take to be a preference rather than an inflexible rule. (Note that the relevant notion of timbre may be somewhat abstract, with the orchestra or a group of winds counting as one timbre rather than many, simply because their component parts are hard to individuate.)

(34) If two musical events involve different timbres, if possible, do not treat them as being coreferential.

Importantly, the converse does not hold: the same timbre can be associated with different objects, as will shortly become clear in a piano piece. Still, when several timbres are used simultaneously, the re-appearance of one timbre may be indicative of coindexing.

4.3   Revealing discourse referents: an example

To highlight the importance of discourse referents in music semantics, we will investigate a Chopin mazurka[23] in which the music without the dynamics is compatible with different coreferential relations among virtual sources. Chopin's dynamics makes some implausible, and we will see that orchestrations (i.e. adaptations of the mazurka for an orchestra) typically make more explicit choices because different timbres are strongly associated with different objects, as stated in (34). Finally, the importance of cross-reference relations will become even more salient when consider a ballet (by Fokine) written for an orchestration of Chopin's music. (In the next section, we will consider recomposed versions of one of these orchestrations so as to evoke very different coreference relations.)

4.3.1   Piano version

We consider the beginning of Chopin's Mazurka Op. 33 No 2, reproduced in (35). It has an extremely simple structure of the form AB, repeated, i.e. AB A'B' with A' = A and B' = B. This is part of a broader structure depicted in (36): the initial pattern AB$_f$ A'B'$_{pp}$ (with the dynamics forte -> pianissimo, notated as f vs. pp) is followed by a different but structurally analogous pattern CD$_f$ C'D'$_{pp}$, before AB$_{ff}$ A'B'$_{pp}$ recurs, but louder (AB is realized fortissimo [= ff] rather than forte [= f], as at the beginning).

---

[23] Mazurkas are traditional Polish folk dances. Chopin's piano pieces by the same name are inspired by them but form a separate genre.

(35) Chopin, Mazurka Op. 33 No 2, beginning
Full score: https://drive.google.com/file/d/15Vz7lKgRDKWawUTxW9ZKWsmoaDT912A8/view?usp=sharing

(36) Broader environment
Full score: https://drive.google.com/file/d/15Vz7lKgRDKWawUTxW9ZKWsmoaDT912A8/view?usp=sharing
$[AB]_f [A'B']_{pp}$
$[CD]_f [C'D']_{pp}$
$[AB]_{ff} [A'B']_{pp}$

We start from a flat realization with constant loudness, as in (37)a. If we assume that there are two virtual sources (which need not be the case), there seem to be at least two salient possibilities: one is that a virtual source corresponds to [AB], while the other corresponds to [A'B']. Alternatively, one virtual source could correspond to A and A', and the other one to B and B'. When we add the dynamics, as in Chopin's score, the second possibility becomes far less likely. The reason is that [AB] is played forte, while [A'B'] is played pianissimo. This can be made sense of if one source is energetic or close and corresponds to [AB], while the other is less energetic or further away and corresponds to [A'B']. Note that nothing blocks an analysis in which a single source is responsible for the entire excerpt, but it gets further away, or loses energy, or intentionally repeats itself with less assertiveness in [A'B']. In addition, one can create an 'anti-Chopin' dynamics, in which A and A' are realized forte, while B and B' are realized pianissimo: if anything, this suggests that A and A' correspond to one source, while B and B' correspond to another (as we explain in Section 4.5, the anti-Chopin dynamics is particularly important because it evokes a coreference between A and A', one that could not be handled in syntactic terms by treating these two components as part of the same group).

(37) a. Flat realization   [AB] [A'B']   https://www.dropbox.com/s/xbmlw1nszjwok1b/Chopin-Mazurka%2033-2%28sans%20nuances%29.aiff?dl=0
      b. Chopin's dynamics $[AB]_f [A'B']_{pp}$ https://www.dropbox.com/s/qq9c61eoaviafxv/Chopin-Mazurka%2033-2%28original%29.aiff?dl=0
      c. Anti-Chopin dynamics $A_f B_{pp} A'_f B'_{pp}$ https://www.dropbox.com/s/vu8tzaugin7wk2l/Chopin-Mazurka%2033-2%28nuances%20inversees%29.aiff?dl=0

As one might expect, most real piano performances follow Chopin's dynamics and thus suggest that, if two sources are present, one corresponds to AB while another corresponds to A'B' (while probably leaving open the possibility that there is a single source present). An example is Rubinstein's rendition in (38)a. But there are also performances that do not follow Chopin's dynamics. Guller's interpretation in (38)b does not draw a clear distinction between AB and A'B' in terms of loudness, which makes it difficult to hear the piece as involving an echo or as one source replying to another.

In terms of the formalism introduced in (31), this means that contraindexing AB and A'B' is compatible with Rubinstein's interpretation (though not forced by it, as A'B' could involve the same source as AB), but not very compatible with Guller's interpretation (the identity of realization suggests that the same source is involved); this is stated in (39).

(38) Two interpretations of the beginning of Chopin's Mazurka Op. 33 No 2
    a. <u>Arthur Rubinstein:</u>[24] $[AB]_f\,[A'B']_{pp}$ https://www.youtube.com/watch?v=P7LDXdjfO5o
    b. <u>Youra Guller:</u>[25] $[AB]_{mf}\,[A'B']_{mf}$ https://www.youtube.com/watch?v=lIIgGWkBx7o&t=551s

(39) $[AB]_1\,[A'B']_2$ with $s(1) \neq s(2)$ is
    a. compatible with Rubinstein's interpretation in (38)a
    b. not very compatible with Guller's interpretation in (38)b
    c. even less compatible with the anti-Chopin dynamics in (37)c (which is suggestive of $[A_1B_2]\,[A'_1B'_2]$, with $s(1) \neq s(2)$).

Our anti-Chopin dynamics in (37)c tends to suggest that, if two virtual sources are involved, the correspond to A and A' on the one hand and B and B' on the other,

(40) Coreference relations evoked by the anti-Chopin dynamics in (37)c
    $A_1B_2\,A'_1B'_2$ with $s(1) \neq s(2)$

To facilitate discussion, we will henceforth assume that when distinct indices are used, their denotations are different as well, hence we won't write things like $s(1) \neq s(2)$ (if we take that difference in denotation is just a possibility, we'll make a note to that effect).

4.3.2   *Orchestrations*[26]

When orchestrating a piano piece, a composer must decide how to associate musical elements with different instruments and timbres. This leads to choices that can disambiguate coreferential relations among virtual sources (because of the principle stated in (34)). We will now consider different orchestrations of the Chopin piece that illustrate this point. We will discuss not just (35) but also the broader environment displayed in (36).

In an orchestration due to Benjamin Britten, the division of labor between instruments is consonant with the division among virtual sources we posited on the basis of Chopin's dynamics. Specifically, the AB sequence played forte in Chopin's score is taken over by the orchestra, whereas the pianissimo A'B' sequence is played by the winds (oboe and flutes), as seen in (41) (the orchestra is primarily made of strings).

(41) **<u>Benjamin Britten's orchestration</u> (1941)**[27] https://www.youtube.com/watch?v=DQLde6QJXvM
    a. $[AB]_{orchestra}\,[A'B']_{oboe+flute}$
    b. $[CD]_{orchestra}\,[C'D']_{oboe+flute}$
    c. $[AB]_{orchestra}\,[A'B']_{oboe+flute}$

While the Chopin dynamics was consonant with the contraindexing in (39), it did not force it because positing a single source was compatible with the difference in loudness between AB and A'B' (in case that source moved away, lost energy or repeated itself less assertively). By contrast, the association between different timbres and different objects is strong enough (as stated in (34)) that the Britten orchestration strongly suggests contradindexing as in (39), expanded as in (42). In view of the

---

[24] Retrieved from https://www.youtube.com/watch?v=P7LDXdjfO5o on December 13, 2019. Credits: Album Artur Rubinstein Plays Chopin Licensed to YouTube by SME (on behalf of RCA Classics).

[25] Retrieved from https://www.youtube.com/watch?v=lIIgGWkBx7o on December 13, 2019; the site gives the information: Youra Guller (1895-1981), recorded 1956.

[26] Finding the origin of different orchestrations is often difficult because information provided with online recordings may be insufficient or even misleading (in case different parts of a ballet music were orchestrated by different composers). The contrasts that we discussed are clear enough that they could be assessed 'by ear', and major differences in instrumentation were investigated (also by ear) by A. Bonetto.

[27] Retrieved from https://www.youtube.com/watch?v=DQLde6QJXvM on December 13, 2019 Credits: Chopin: Les Sylphides (Arr. for Orchestra By Benjamin Britten, Mono Version) Joseph Levine, American Ballet Theatre Orchestra 1 January 1954. For the history of the Britten score (lost and then rediscovered), see Cooper 2013.

identity of timbre between CD and AB, and also between A'B' and C'D', we take this orchestration to make it plausible that only two sources are present, written as 1 and 2 in (42).[28]

(42) **Coreference relations suggested by Britten's orchestration**
     a. $[AB]_1 [A'B']_2$
     b. $[CD]_1 [C'D']_2$
     c. $[AB]_1 [A'B']_2$

     A rather different choice is made by another orchestration (whose author we have not yet identified), which contrasts the orchestra (used for AB, CD and again AB) with two groups of winds : one, for both occurrences of A'B', includes the oboes (= winds1); the other, for C'D', includes the clarinets (= winds2), as is represented in (43). This suggests the coreference relations depicted in (44).

(43) **Other orchestration**   https://www.youtube.com/watch?v=9cb6ou3Kklg&t=11m35s
     a. $[AB]_{orchestra} [A'B']_{winds1}$
     b. $[CD]_{orchestra} [C'D']_{winds2}$
     c. $[AB]_{orchestra} [A'B']_{winds1}$

(44) a. $[AB]_1 [A'B']_2$
     b. $[CD]_1 [C'D']_3$
     c. $[AB]_1 [A'B']_2$

     A more complex pattern of indexing is suggested by Roy Douglas's orchestration in (45): while the beginning contrasts the orchestra with the winds, each new appearance of the winds (A'B', C'D' and A'B' again) features distinct groups of instruments, three in total. This invites an interpretation whereby a source 1 corresponding to the orchestra interacts with three distinct sources 2, 3, 4, as summarized in (46) (we disregard the fact that the orchestra takes slightly different forms in (45)a, b and c).

(45) **Roy Douglas's orchestration**[29] https://www.youtube.com/watch?v=XpwGTY9T5UE
     a. $[AB]_{orchestra} [A'B']_{winds1\ oboe\ bassoon}$

---

[28] Broadly similar choices are made in an orchestration by Gordon Jacob. The timbres are those in (i), where the orchestra is clearly contrasted with the winds, while the identity of the winds is a bit underspecified. To Bonetto's ear, winds1 include flutes or piccolos and a bassoon, winds1? flutes or piccolos and something else; the distinction is less than obvious, so relations of coreference should be encoded with some uncertainty as in (ii) (hence 2? on [C'D']).

(i)   **Gordon Jacob's orchestration**   https://www.youtube.com/watch?v=rGi_mW_elkA&t=7s
     a. $[AB]_{orchestra} [A'B']_{winds1}$
     b. $[CD]_{orchestra} [C'D']_{winds1?}$
     c. $[AB]_{orchestra} [A'B']_{wind1}$

(ii)   a. $[AB]_1 [A'B']_2$
     b. $[CD]_1 [C'D']_{2?}$
     c. $[AB]_1 [A'B']_2$

     Maurice Keller's orchestration in (iii) also contrasts the orchestra for AB with a group of winds for A'B' (= winds1), but then things become less clear. The orchestra recurs for CD, but for the second AB, it appears in modified form (orchestra+) together with a countermelody by the cellos, which we disregard here (it could be treated as a separate source). The winds appear in modified form for C'D' (written as winds1+), and a version of the winds appears again for the last A'B'. This yields the same type of indexing as the Britten orchestration, but with much greater uncertainty, as shown in (iv) (where the indices followed by ? may but need not mark coreference).

(iii)   **Maurice Keller's orchestration** (1908) https://www.dropbox.com/s/r5bixj70sry9b97/Chopin-Mazurka%20Op.33%20No2-Keller.mp4?dl=0
     a. $[AB]_{orchestra} [A'B']_{winds1}$
     b. $[CD]_{orchestra} [C'D']_{winds1+}$
     c. $[AB]_{orchestra+} [A'B']_{winds1+}$

(iv)   a. $[AB]_1 [A'B']_2$
     b. $[CD]_1 [C'D']_{2?}$
     c. $[AB]_{1?} [A'B']_{2?}$

[29] Retrieved from https://www.youtube.com/watch?v=jSQbDXEGNUA on December 13, 2019. Credits: Heinz Fricke Album Delibes, L.: Coppelia Ballet Suite / Chopin, F.: Les Sylphides (orch. R. Douglas) (Berlin Radio Symphony, Fricke)

    b. $[CD]_{orchestra} [C'D']_{winds2\ flute}$

    c. $[AB]_{orchestra} [A'B']_{winds3\ clarinet}$

(46) a. $[AB]_1 [A'B']_2$

     b. $[CD]_1 [C'D']_3$

     c. $[AB]_1 [A'B']_4$

       Finally, in an orchestration similar to or identical to one by Arthur Fiedler, the orchestra gets enriched as it transitions from A to B and again as it transitions from C to D, as is represented in (47)a,b (where *orchestra+* refers to the enriched orchestra), before the enriched orchestra is used for the entire AB segment in (47)b. A'B', C'D' and the second A'B' involve three different groups of instruments: winds1 (including the oboes) for the first A'B', winds2 (including the flutes) for C'D', and strings for the second A'B'. Making use of the notation for sums of indices in (33), we can represent the coreferential relations suggested by this orchestration in (48), where there is overlapping reference among various sources (whether 4 should be taken as a version of the unenriched orchestra isn't entirely clear).[30]

(47) **Similar to Arthur Fiedler**[31]   https://www.youtube.com/watch?v=T46L0brOu_0

     a. $[A_{orchestra} B_{orchestra+}] [A'B']_{winds1}$

     b. $[C_{orchestra} D_{orchestra+}] [C'D']_{winds2}$

     c. $[AB]_{orchestra+} [A'B']_{strings}$

(48) a. $[A_1 B_{1+1'}] [A'B']_2$

     b. $[C_1 D_{1+1'}] [C'D']_3$

     c. $[AB]_{1+1'} [A'B']_{4\ (or\ 1?)}$

       Two conclusions can be drawn. First, through the use of timbre (and the preference rule in (34)), orchestrated music can make coreference relations clearer than piano music. Second, different orchestrations can make difference choices, but not anything goes: Chopin's dynamics (in (39)) suggested that AB and A'B' correspond to different virtual sources, and all the orchestrations discussed here make similar choices in this respect.

4.3.3   *Ballet*

We now further highlight the importance of coreference relations by discussing their role in a ballet set to an orchestrated version of Chopin's music.

       When creating a ballet for a music, somewhat similar issues arise as in orchestration because different dancers may be associated with different parts of the music. In general, the relation between dance and music need not be a simple one: each medium arguably has its own abstract semantics, and while there must be points of contact so that a sense of unity is gained, there is no requirement that the dance should convey the same information as the music, just in a different modality.[32] Still, by focusing on particularly strong points of contact between the music and the dance, one can attempt to use the latter to help reveal semantic aspects of the former.

       A simple case is offered by Michel Fokine's ballet *Les Sylphides* (originally called *Chopiniana)*, which contains a part on an orchestrated version of the Chopin mazurka discussed above (we believe the original version was created for Maurice Keller's orchestration – see fn. 28). Strikingly, AB in (36) corresponds to a movement by the main ballerina, while A'B' corresponds to a movement by the other dancers, as seen in (49), with a similar division in CD (main ballerina) vs. C'D' (other dancers), before the main ballerina appears again for the final AB, and the other dancers, now joined by the main ballerina, dance the last A'B'. This is indicative of a pattern of indexation that is almost but not quite identical to that of the Britten orchestration (as well as to those of Gordon Jacob and Maurice Keller

---

[30] We disregard a further complexity (which isn't so easy to perceive): winds1 get slightly modified between A' and B', and similarly for winds2 between C' and D'.

[31] An orchestration explicitly attributed to Arthur Fiedler is very similar but slightly harder to hear (https://www.youtube.com/watch?v=k1nnNthX4eU).

[32] The same issue arises with film and cartoon music: it may complement rather than repeat the content of the visual scenes and dialogues (specialists use the term 'mickey mousing' when the music conveys the same information as the visuals, and this need not be laudatory).

– see fn. 28): the main ballerina corresponds to source 1, the other dancers (treated as a unique character due to their unity of movement) to source 2. The only difference from the Britten orchestration is that the final A'B' does not just correspond to 2 but to 2 joined by 1, as is represented in (50).

(49) Fokine's *Les Sylphides* (originally called *Chopiniana)*, movement on Chopin's Mazurka Op. 33 No 2[33]
   [Performance from 1984](https://www.youtube.com/watch?v=LBJNc3h7Hp8&t=10m46s), American Ballet Theatre, on an orchestration close to Britten's version
   https://www.youtube.com/watch?v=LBJNc3h7Hp8&t=10m46s
   $[AB]_{\text{main ballerina}} [A'B']_{\text{other dancers}}$
   $[CD]_{\text{main ballerina}} [C'D']_{\text{other dancers}}$
   $[AB]_{\text{main ballerina}} [A'B']_{\text{other dancers(+main ballerina)}}$

(50) a. $[AB]_1 [A'B']_2$
      b. $[CD]_1 [C'D']_2$
      c. $[AB]_1 [A'B']_{2+1}$

In sum, a source-based semantics naturally gives rise to the issue of cross-reference among different sources. While the music may be underspecified in this respect, especially when written for a single instrument, properties such as loudness may favor some patterns of coindexing. These can be brought out more clearly by orchestration because different timbres are naturally associated with different sources. Dance can further highlight the importance of cross-reference by making concrete the identification of certain dancers with certain sources, hence strongly favoring some patterns of cross-reference over others.

4.4 *Creating minimal pairs*

For the analysis of a piece to be convincing, it is not enough to show that some parameters *might* explain some semantic impressions. One needs to show that when these parameters are minimally modified, the semantic impressions change accordingly.

We already saw in (37)c and (40) that inversing Chopin's dynamics can evoke very different patterns of cross-reference. But the point can be made more strongly by modifying the orchestration, as different timbres are particularly strongly associated with different sources (as stated in (34)). We start from a highly simplified version of Britten's orchestration, in (51)a, with AB played by the orchestra and A'B' by the flute and oboe. Our initial version is based on the piano score, but with the right hand (i.e. the part corresponding to the melodic theme) replaced by an orchestra timbre for AB, and by a flute and oboe timbre for A'B'. We then modify the assignment of timbres: in (51)b, A is played by the orchestra and B' by the flute and oboe, and then A' is again played by the orchestra while B' is played by the flute and oboe.

(51) a. [Simplified version of Britten's orchestration](https://www.dropbox.com/s/7mamktvqerdzuee/Chopin-Mazurka%2033-2%28cc_vv%29.aiff?dl=0) https://www.dropbox.com/s/7mamktvqerdzuee/Chopin-Mazurka%2033-2%28cc_vv%29.aiff?dl=0

Orchestra	A - B
Oboe+flute	A' - B'

   b. [Anti-Britten](https://www.dropbox.com/s/ew2xgtq2k7zgnmw/Chopin-Mazurka%2033-2%28cv_cv%29.aiff?dl=0) (A. Bonetto) https://www.dropbox.com/s/ew2xgtq2k7zgnmw/Chopin-Mazurka%2033-2%28cv_cv%29.aiff?dl=0

Orchestra	A, then A'
Oboe+flute	B, then B'

The result is striking: the 'anti-Britten' orchestration in (51)b gives the impression of a dialogue between two entities: A gives rise to the reply in A', B gives rise to the reply in B'. This achieves by way of timbres, and in a stronger way, the coreferential relations evoked by the anti-Chopin dynamics in (37)c.

Once it is made plausible that discontinuous components can be given an identity by way of a coreference relation, we can go one step further and add properties to the music to further characterize the objects involved. Lerdahl 2001 proposed this idea when he compared musical events to Heider and Simmel's (1944) abstract animations "in which three dots moved so that they did not blindly follow physical laws, like balls on a billiard table, but seemed to interact with another – trying, helping, hindering, chasing – in ways that violated intuitive physics", and thus were perceived as animate agents

---

[33] Several sources suggest that this piece was added to Fokine's original version of *Chopiniana* on an orchestration by Maurice Keller. Thus Craine and Mackrell 2010 (p. 435) imply that Maurice Keller orchestrated the additional pieces (thus the Mazurka Op. 33 No. 2) that were added to the original ballet in the version premiered on March 21, 1908 at the Mariinsky theater in St Petersburg, Russia.

(video examples: http://bit.ly/2CR5AB2). Lerdahl argued that similar effects arise in music: "here the dots are events, which behave like interacting agents that move and swerve in time and space, attracting and repelling, tensing and coming to rest".[34]

      Despite the highly abstract character of music semantics (hence the fact that Bernstein's Don Quixote and Superman interpretations are just two possible stories among many), one can easily modulate the voices so as to evoke different properties of the sources. A radical example was given in (8) and (9): reversing the melodic direction of the tune corresponding to Don Quixote's departure radically altered the character we attributed to him. More subtle modifications can be made to the recomposed piece in (51)b. While keeping the coreferential relations constant (with one source corresponding to A and A', and another to B and B'), we will modulate the tempo, loudness and articulation to modify the character attributed to each source. In (52)a, AA' appears as far more assertive than BB'; the reason is that in AA' is the music is faster, louder and more accented (staccato). In (52)b, BB' is played faster, louder and more staccato, and seems more assertive as a result.

(52)   a. <u>AA' assertive - BB' fearful</u> (A. Bonetto) https://www.dropbox.com/s/dno710tcvs4jnjc/Chopin-Mazurka%2033-2%28assertif_craintif%29.mp3?dl=0
      AA' (and the piano) is faster, louder, more accented (staccato).
      b. <u>AA' fearful - BB' assertive</u> (A. Bonetto) https://www.dropbox.com/s/r7wmzu3930wd4ef/Chopin-Mazurka%2033-2%28craintif_assertif%29.mp3?dl=0
      BB' (and the piano) is faster, louder, more accented (staccato).

      In sum, the anti-Britten (and anti-Chopin) patterns of coreference explored in (51) (and (37)c) confirm a point already highlighted by the Don Quixote excerpts in (30): coreference is not reducible to syntactic notions such as grouping. A more abstract semantic notion is needed, corresponding to identity of virtual sources. Once this cross-referential device is in place, the sources can be endowed with further individual properties depending on nature and interpretation of the music.

4.5   *Syntactic vs. semantic analyses*

The minimal pairs created in Section 4.4 make an important theoretical point. When analyzing Chopin's dynamic for $[AB]_f\ [A'B']_{pp}$ and the corresponding (real-life) orchestrations, we could have captured the main intuitions in a syntactic fashion, by using the notion of 'groups' from Lerdahl and Jackendoff's work on musical syntax. On this view, A and B in AB form a natural unit not because they correspond to the same virtual source, but because they form a syntactic group, and similarly for A' and B' in A'B' (in Lerdahl and Jackendoff's theory, what matters at this point is what they call 'grouping structure'[35]). This syntactic explanation was not available in our initial analysis of the reappearance of the Don Quixote virtual source in (30) because the excerpts appeared at a considerable distance, making it impossible to treat them as a natural syntactic unit. The recomposed music in (37)c and especially (51)b makes the same point in a more minimal fashion: by making choices of timbre that strongly suggest that A and A' correspond to the same virtual source despite the intervening B, the recomposed version shows that that syntactic grouping cannot explain this connection. A semantic analysis, by contrast, can easily provide it.

      Of course one cannot entirely exclude the possibility that some *other* syntactic notion might treat A and A' as a natural unit in this case (and similarly for B and B'). For instance, as suggested by E. Chemla (p.c.), one could note that A and A' play symmetric roles in two different groups. But what it striking is that this structural fact holds irrespective of the dynamics and timbre, and thus the latter are crucial in giving A and A' a joint identity or not: a structural relationship seems to be insufficient to account for the effects we observed.

---

[34] See Lerdahl 2019, Chapter 3, for further remarks about the relevance of Heider and Simmel's animations to music semantics.

[35] As summarized by Lerdahl 2001, "grouping structure describes the listener's segmentation of the music into units such as motives, phrases, and sections"; it is explicitly derived in Lerdahl and Jackendoff 1983 from Gestalt principles of perception. See Schlenker 2017, 2019a for an attempt to reinterpret grouping in terms of the mereological structure of the denoted events

4.6 Intermediate conclusion and refinements

If the present analysis is on the right track, then, a performer and a listener must typically make decisions about patterns of cross-reference among virtual sources. These are not reducible to syntactic relations of grouping, which involve contiguity. In music written for a single instrument, such as the piano, coreference relations are typically underspecified, although they are constrained by dynamics and further properties of the music and interpretation. Orchestration typically makes coreference relations more explicit because difference of timbre is strongly associated with disjoint reference. When music is combined with dance and the two are understood to be in strong correspondence, coreference relations become even more salient. Finally, once such relations are established in a music excerpt, they can also be combined with semantically important musical modulations so as to convey different kinds of information about the virtual sources, which may be animate or inanimate, assertive or fearful, etc. While the semantics of music is typically far more abstract than that of visual narratives, the notion of coreference introduced by Abusch in the latter can illuminate the former as well.

5   Cosuppositions triggered by music I: co-speech and co-film/gif music

We turn to a different aspect of musical meaning that can benefit from other areas of Super Semantics. In brief, we will argue that, in some cases at least, music can affect the meaning of film in the same way as gestures and facial expressions affect the meaning of speech.

This may seem like a far-fetched idea in view of the difference between the relevant mediums. But there is one crucial similarity between co-speech gestures (or facial expressions) and co-film music: both are typically taken to be parasitic on the message they enrich (and it is only to the extent that this is in fact the case that we expect to find precise similarities between the two cases). Specifically, in standard cases a fully well-formed and comprehensible message is preserved when a co-speech gesture is disregarded. Similarly for film music: the heart of the action is usually carried by the visuals and dialogues, and the music can in this sense be taken to be parasitic on them (which need not mean that it is unimportant). It was argued in earlier work that, possibly due to their own parasitic status, co-speech gestures or facial expressions trigger a special kind of presupposition, called *cosupposition* (Schlenker 2018a,b). Pasternak 2019 extended the same generalization to co-speech sound effects (implicitly with an eye to co-speech music). We will argue that film and cartoon can music trigger cosuppositions as well.

5.1   Gestural cosuppositions

While the formal semantic analysis of co-speech gestures and facial expressions is still in its infancy, we will follow recent analyses that argue that some of these expressions trigger conditionalized presuppositions (but see Lascarides and Stone 2009 and Ebert and Ebert 2014 for rather different views). Schlenker 2018a,b focuses on gestures that co-occur with propositional and predicative expressions and argues that examples such as (53)a(i) differ from at-issue controls like (53)(ii) in triggering a presupposition. The presupposition is conditionalized on the contribution of the modified expression, and in the present case it has the form: *if x helps x's son, lifting will be involved;* it is because of this conditionalization that this presupposition is called a *cosupposition*.[36] Questions are an environment in which presuppositions project, as seen in (53)b. The important observation is that the conditionalized inference *if John helps his son, lifting will be involved* projects in the same way in (53)(i) (but not in the control in (53)(ii)).

Notation: gestures are encoded `in special font` (sometimes with a picture) *before* the expressions they occur with, which is **boldfaced**.

---

[36] As noted in Schlenker 2018a, "the terminology is intended to suggest that a cosupposition triggered in a local context c' is computed in tandem with ('co') an at-issue component in c' (by contrast, a standard presupposition triggered in c' is computed before ('pre') any at-issue component in c')".

(53) a. Will John (i) UP_ **help** his son? (ii) help his son like UP_ **this**?

(i) => If John helps his son, he will do so by lifting him

b. Will John realize that his son is losing?
=> John's son is/will be losing

The presuppositional nature of the conditionalized inference can be assessed in multiple other environments, such as under *none*-type quantifiers, as in (54):

(54) a. None of these 10 guys (i) UP **helped** his son. (ii) helped his son like UP **this**.

(i) => none of these 10 guys helped his son; but for each of them, if he had helped his son, it would have been by lifting him
b. None of these 10 guys realized that his son needed help.
=> for each of these 10 guys, his son needed help

Co-speech facial expressions appear to trigger cosuppositions as well, and certain non-grammatical facial expressions in ASL (American Sign Language) arguably behave in the same way (Schlenker 2018a, b).
Notation: :-(is a disgusted facial expression co-occurring with the boldfaced expression.

(55) a. Did Sam go :-(**[skiing with his parents]**?

=> for Sam to go skiing with Sam's parents would be disgusting (from Sam's / from the speaker's standpoint)
b. None of my friends goes :-(**[skiing with his parents]**.

=> for each of my friends, to go skiing with his/her parents would be disgusting (from the friend's / from the speaker's standpoint)

Tieu et al. 2017, 2018 provide experimental evidence that conditionalized inferences triggered by co-speech gestures broadly project like presuppositions.

5.2   The generality of cosuppositions

Cosuppositional inferences have been argued to arise with non-gestural material as well. Suppose I am commenting, in writing, on the French comic *Asterix*. I could ask at some point:

(56)

What will happen next: will Asterix …   **…do what's needed?**

Although the target clause is a question, there is a strong intuition that *what is needed* involves drinking the magic potion. In other words, one obtains a cosupposition of the form: *if Asterix does what's needed, drinking the magic potion will be involved*.

Closer to our topic, Pasternak 2019 argues that sound effects that co-occur with words trigger cosuppositions as well. In (57)a, the sound of an explosion co-occurs with the Verb Phrase, and triggers the inference that if the soldier were to assassinate his target, an explosion would be involved; no such inference arises with the at-issue control in (57)b.[37]

(57) a. The soldier will not BOOM [assassinate his target]. https://www.dropbox.com/s/za906j1bqz6coan/Pasternak-explosion-neg.wav?dl=0
=> if the soldier were to assassinate his target, an explosion would be involved
b. The soldier will not assassinate his target like BOOM this. https://www.dropbox.com/s/z6ssjygvu2ou5fx/Pasternak-explosion-neg-control.wav?dl=0

---

[37] Thanks to Rob Pasternak for authorizing us to cite his sound files.

≠> if the soldier were to assassinate his target, an explosion would be involved
(Pasternak 2019)

We should add that Pasternak was initially interested in musical effects co-occurring with speech, and some of his examples are indeed musical in nature, as in (58), where a descending scale co-occurring with the Verb Phrase suggests that if the student werer to adjust the brightness, this would involve turning it down.

(58) **The student will not DOWN [adjust the brightness setting of his computer screen]**.
https://www.dropbox.com/s/oua541484ycx80j/Pasternak-piano-neg-down.wav?dl=0
=> if the student were to adjust the brightness, this would involve turning it down

Pasternakian examples can be produced with real music as well. In (59), the beginning of Rossini's William Tell overture is superimposed on the Verb Phrase. The latter alone would be somewhat underspecified, but with the music one gets the clear sense that if the cavalry does what's needed, it will come quickly and triumphantly, two properties that are evoked by the music on its own.

(59) [Phlegmatic pianist, to the mayor of a besieged city]
Sir, I am told the enemy is about to enter the city. Will our cavalry [MUSIC] **[do precisely what's needed at the present moment]**?
Music alone: https://www.dropbox.com/s/vbsksds7y0wnqqb/Rossini-Tell%20original-louder.mp4?dl=0
Speech+music: https://www.dropbox.com/s/51k3927fupgdrxt/Rossini-Tell-co-speech%20music.wav?dl=0
=> if the cavalry does what's needed, it will come quickly and triumphantly

Why do cosuppositions arise with such diverse means of enrichment? While a full derivation has yet to be given, it was speculated (e.g. Schlenker 2018a,b, to appear) that an enrichment p' that shares a time slot with the main message p should be such that p' can be disregarded without affecting the meaning of p relative to its context c'. The relevant notion of context is that of a 'local context' as used in presupposition theory (e.g. Heim 1983, Schlenker 2009, 2010, 2011). The result is that, relative to c', pp' (i.e. the conjunction of p and p') should be equivalent to p, or in other words: p should entail p', as is stated in (60) (see Esipova 2019 for a slightly more general statement).

(60) a. A cosupposition is triggered when an elementary expression (possibly including a co-speech/sign gesture) pp' has an entailment p' which is presented as being unimportant, and for this reason the global Context Set C should guarantee that, relative to its local context c', pp' should be equivalent to p, i.e.

(i) c' |= pp' <=> p

b. (i) is equivalent to the standard definition of cosuppositions in (ii):

(ii) c' |= p => p'
(Schlenker, to appear)

If these intuitions are on the right track, one expects to find cosuppositions in other cases in which an enrichment is somehow secondary relative to the main message. We will now argue that this in fact happens with film and cartoon music.

5.3   *Co-film music can trigger cosuppositions*

Let us start with an example. In Kubrick's *2001: A Space Odyssey*, a key moment towards the beginning involves an ape playing with and then destroying some bones.
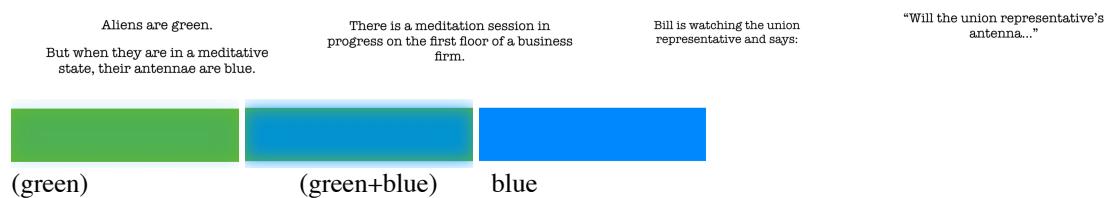
(61) Kubrick's bones scene (from *2001: A Space Odyssey*) https://www.dropbox.com/s/c8ggdz5ls4dtmjw/Bones-Original-long.mp4?dl=0

This turns out to represent the invention of tools, hence a pivotal moment in human evolution. While the end of the visuals is rather explicit in this respect, the beginning of the scene seems innocuous enough: it just displays an ape playing with some bones. But Kubrick superimposed on the visuals the music from Strauss's Zarathustra, intended to evoke a sunrise (and in fact the opening of the movie features the beginning of the Strauss piece with a planet appearing behind another planet). The end of the Kubrick scene acquires a more momentous meaning thanks to the music. But the music also completely changes the meaning of the *beginning* of the scene: it indicates that the apparently innocuous actions are in fact momentous.

While it is enough to compare the original version with the muted version to see that the music contributes something to the meaning of the scene (more specifically, it infuses it with a momentous or triumphant meaning), this tells us nothing about the status of this inference: is it at-issue? cosuppositional? something else? Or are these distinctions meaningless for a film? We can begin to address the question, by inserting the scene within a sentence. This may seem like an odd idea, but the investigation of gestures and even visual animations that entirely replace some words ('pro-speech[38] gestures and visual animations') has recently yielded some insights about the inferential typology of language. Specifically, it was argued in Schlenker 2019c that pro-speech gestures trigger a variety of inferential types, including presuppositions, which can be empirically characterized by their interaction with various logical operators. Tieu et al. 2019 confirmed a subset of the generalizations with experimental means and extended them to pro-speech visual animations.[39] For instance, in (62), it was stated that aliens are green, but that when they are in a meditative state, their antennas are blue. The experiment was designed to show that an animation depicting an alien's antenna turning from green to blue triggers a *presupposition* that the alien is not initially meditating.

(62) **Pictures from Tieu et al.'s videos testing presuppositions generated by visual animations**
(here: a change of state animation pertaining to an alien's antenna turning from green to blue; original video: https://youtu.be/U6dfs-XI2-4

Aliens are green. But when they are in a meditative state, their antennae are blue.	There is a meditation session in progress on the first floor of a business firm.	Bill is watching the union representative and says:	"Will the union representative's antenna…"
(green)	(green+blue)	blue	

Instead of pro-speech visual animations, we will use pro-speech film excerpts and gifs, but we will accompany them with music to test the status of the semantic enrichment contributed by the latter. A first stab is provided in (63). The more innocuous part of the Kubrick visuals, notated as *APE-BREAKING-BONES,* is embedded in a discourse, while being accompanied by Strauss's music, notated as a superscript (Zarathustra). The part written between curly brackets appears in the video in a bubble to indicate that this is imagined (without it, the fact that the scene is so precisely depicted tended to trigger the inference that it was presented as real).

(63) Embedding a Kubrick excerpt in a sentence https://www.dropbox.com/s/y43x76j0p3lk9pt/Bones-Question-Strauss-Bubble.mp4?dl=0
I saw an ape playing with some bones…
And I wondered…
{Will it… Zarathustra**APE-BREAKING-BONES**}

It is clear that the sentence with the embedded video excerpt is interpreted as something like: *Will it break the bones?* By comparing the muted to the full version, we can see that, in this case at least, whatever inferences were contributed by the music in (61) are not at-issue, and are arguably conditionalized in nature. So one would be inclined to endorse the inferences in (64)d and possibly (64)c,e, but certainly not (64)a,b.

(64) **Possible cosuppositional inferences**
If the ape breaks bones, this will be …
a. a terrible thing
b. something light-hearted
c. something positive
d. an accomplishment
e. a fateful action

This initial attempt to assess the status of the musical enrichment is imperfect, however. For starters, the excerpt is too short to make it possible for the music to form a coherent whole, as it was

---

[38] Just like a co-speech element accompanies a linguistic expression, a pro-speech element *replaces* one.
[39] Guerrini and Schlenker 2019 and Guerrini and Migotti 2019 further extend these claims to pro-speech onomatopoeias and pro-speech music, a point to which we briefly return below.

inserted in the movie so as to have its initial climax a bit later in the film, when the tools are used to kill animals. In addition, nothing decisively tells us at this point that it is the precise content of the music (rather than the presence of music in general) which is responsible for the inference. To address these objections, we turn to artificially modified versions of the excerpt, with other famous tunes replacing the Strauss music; in effect, we are using the method of minimal pairs to assess the important of film music. The examples in (65) were constructed so as to offer a better coordination between the music and the film excerpt.

(65) **Modifyng the Kubrik excerpt, with different musics**
I saw an ape playing with some bones…
And I wondered…
{Will it…
a. ^{Carmen}**APE-BREAKING-BONES**}   video: [Carmen](...) Prelude to Act I (Leonard Slatkin)
https://www.dropbox.com/s/z4p8irsxo4bu57s/Bones-Question-Carmen%2B%2B-Bubble.mp4?dl=0
b. ^{Beethoven 8th}**APE-BREAKING-BONES**}   video: [Beethoven 8th](...), last bars (Kurt Masur)
https://www.dropbox.com/s/xm99mj1lyzb4tre/Bones-Question-Beethoven-Bubble.mp4?dl=0
c. ^{Whistle}**APE-BREAKING-BONES**}   video: [Billy Mowbray Uke and Whistle](...)
https://www.dropbox.com/s/h3uyewc7fdbrjqo/Bones-Question-Whistles-Bubble.mp4?dl=0

Our goal is not to explain how the selected excerpts trigger the inferences they do, which are very different from those of Strauss's Zarathustra: the evocation of something sinister and fateful for our Carmen excerpt (= (65)a), of the triumphant attainment of a goal for the last bars of Beethoven's 8th Symphony (= (65)b), and of something light-hearted for our whistle tune (= (65)c). But what is rather clear is these inferences are not targeted by the question. Rather, despite the embedding within a question, one gets conditionalized inferences characteristic of cosuppositions. For the Carmen excerpt, one might thus get the inference that the ape's breaking the bones would be fateful and terrible; for the Beethoven excerpt, that it would be positive and an accomplishment; and for the whistle tune, that it would be light-hearted.

5.4   Refining the argument

Our argument has a flaw, however. The problem stems from the very simplicity of our excerpts: the music entirely co-occurred with the visuals, which makes it possible to develop two analyses. According to the more conservative one, the film excerpt comes (by whatever means) to have a verbal meaning, e.g. 'break the bones' in (65). Then it is this entire Verb Phrase that gets modified as a unit, just as if it were made of words. On this view, the details of the music do not enrich the details of the film. Rather, we just have a more complicated instance of the co-speech music in (58)-(59), with the difference that the verbal element happens to be expressed by a film excerpt. On a more radical theory, the music modifies the details of the film, and in particular there are cosuppositional inferences that get triggered below the level of the film excerpt.

We argue for this second, more radical theory. Intuitively, the excerpts discussed in (65) trigger different inferences about different parts of the action. For instance, the Beethoven piece has its final, conclusive chord aligned with the point at which the ape smashes the skull in front of him. The fact that the end of the Beethoven piece is triumphant and conclusive strongly contributes to the inference that the ape's smashing the skull is the accomplishment of a goal. Still, these observations need to be made sharper. To start doing so, we investigate a gif, and we add music to different parts to assess the specific interaction between the music and the cartoon parts (rather than the global meaning contributed by the cartoon as a whole).

We start from the gif in (66), which displays Asterix the Gaul drinking a magic potion before hitting a Roman soldier and leaving the premises.[40]

---

[40] Original from https://giphy.com/gifs/paf-asterix-ulCTAq0E5ekV2, retrieved on December 10, 2019. We modified the original so as to make the three components easier to separate. Specifically, we made the point at which Asterix pauses after drinking the potion longer, and we made his departure slower.

(66) Three stages of a gif featuring Asterix and a Roman soldier

Beginning: Asterix drinking   Middle: Asterix hitting   End: Asterix leaving

We then enrich this gif with a short or long musical excerpt that either (i) covers the entire sequence starting with the drinking, or (ii) only the part that follows the hitting, as Asterix leaves the premises. Musical excerpts are of three kinds: (a) the light-hearted 'Uke and Whistle' tune used in (65)c (henceforth *WHISTLE*); (b) a happy, triumphant commercial tune labeled 'Vintage News' (= *NEWS*); (c) a commercial tune intended to evoke suspense ('Suspense accents 07'; henceforth *SUSPENSE*). We put the excerpt name in capitals (preceded by ♪) above the part of the gif it co-occurs with, with a line to indicate the extent of the co-occurrence;[41] thus in (67)a, the tune *WHISTLE* starts as Asterix drinks the potion and continues through the end of the gif, while in (67)a' it starts after Asterix has hit the Roman soldier.

(67) *Context:* Asterix had an earlier encounter with a Roman soldier. Now he is faced with him once again.[42]

What will happen next? Will Asterix…

a.   ♪ WHISTLE   https://www.dropbox.com/s/ysx313tuds310j3/Asterix-co-gif-a1.mp4?dl=0

 ?

a'.   ♪ WHISTLE   https://www.dropbox.com/s/cpsgjid59iohofx/Asterix-co-gif-a2.mp4?dl=0

 ?

b.   ♪ NEWS   https://www.dropbox.com/s/ikxop9v5m3d7qat/Asterix-co-gif-b1.mp4?dl=0

 ?

b'.   ♪ NEWS   https://www.dropbox.com/s/ikxdcskvxz1ok6n/Asterix-co-gif-b2.mp4?dl=0

 ?

c.   ♪ SUSPENSE   https://www.dropbox.com/s/4zj76bxesw6os7o/Asterix-co-gif-c1.mp4?dl=0

 ?

c'.   ♪ SUSPENSE   https://www.dropbox.com/s/wu80ypnnw9b1ktr/Asterix-co-gif-c2.mp4?dl=0

 ?

Since the gif is embedded in a question, one does not derive an inference that the scene will in fact take place. But we believe that a cosuppositional inference is nonetheless derived, to the effect that

---

[41] This borrows a notation for non-manual expressions in sign language linguistics.

[42] Credits: ambient sound (all excerpts): iMovie audio library, Ambient effect 2. WHISTLE: Billy Mowbray Uke and Whistle NEWS: iMovie audio library, Vintage news short. SUSPENSE: iMovie audio library, SuSPense accents 07; POISON: Simon Boccanegra, Act II, Scene 8, Teatro La Fenice 2014-2015, conductor Myung-Whun Chung, RAI, with Simone Piazzola as Simon.

if Asterix does X (with X corresponding to the entire scene, or just to the character's departure after his deed), this will have a certain character (such as: light-hearted, triumphant or mysterious).[43]

(68) Inferential questions of the form:

If Asterix does X, this (entire action) will be Y

X1 = drinks the magic potion, hits the Roman soldier and leaves
X2 = leaves after drinking the magic potion and hitting the Roman soldier

Y1 = light-hearted
Y2 = triumphant
Y3 = mysterious

Let's focus in particular on the contrast between (67)a and (67)a'. In the first case, the light-hearted whistle tune co-occurs with Asterix's entire action, suggesting that the whole sequence would light-hearted and possibly routine if it were to happen. In other words, one gets a cosupposition to the effect that *if Asterix drinks the magic potion, hits the Roman soldier and leaves, something light-hearted will be involved throughout that sequence*. In the second case, the tune only co-occurs with the part that follows the violent action, which suggests that having done so would lead to a light-hearted situation – for instance, Asterix might be light-hearted after doing his deed. Here the cosupposition is that *if Asterix drinks the magic potion and hits the Roman soldier, then if he leaves the premises, the latter situation will be light-hearted*. We believe that related contrasts arise with the other tunes: in (67)b, the triumphant tune co-occurs with the entire sequence, and one gets the impression that the entire scene would be triumphant if it were to take place. In (67)b', the triumphant tune only co-occurs with the part in which Asterix leaves the premises, and this suggests that if Asterix drank the potion, hit the Roman soldier and left, the latter action (leaving) would be triumphant. Similarly, in (67)c an air of mystery is conferred by the music to the entire sequence of events, whereas in (67)c' it is Asterix's departure which is somehow mysterious.

5.5 *Analyzing the contrasts*

To understand how the contrasts are derived, we will greatly simplify matters and treat our gif as a narrative sequence made of just two pictures, $<P_1, P_2>$, with Asterix hitting the Roman soldier in P_1 and leaving the room in P_2:

(69) Simplified analysis of the gif (66) in terms of two pictures $<P_1, P_2>$

P_1 P_2

In a more realistic analysis, we would analyze gifs as continuous picture sequences, but we simplify things maximally so as to bring out the predictions of a cosuppositional theory: different cosuppositions are obtained depending on which parts of the gif are accompanied by music. Applying the analysis of narrative sequences in (24) to the case at hand, we obtain the truth condition in (70).

(70) The picture sequence $<P_1, P_2>$ is true of situations $<e_1, e_2>$ relative to viewpoint v along the system of projection S iff
(1) temporally, $e_1 < e_2$,
(2) $proj_S(e_1, v) = P_1$ $proj_S(e_2, v) = P_2$.

We will simplify things even more by considering a pair of situations $<e_1, e_2>$ that is temporally ordered, with $e_1 < e_2$, and by asking under what conditions $<P_1, P_2>$ (viewed as a predicate of pairs of

---

situations) is true of $<e_1, e_2>$. Writing $[[<P_1, P_2>]]^{S, v, w}$ for the denotation of $<P_1, P_2>$ in the world w relative to the viewpoint v in the system of projection S, we can write the condition in (70) in the more familiar semantic format in (71).

(71) Let S be a system of projection and v a viewpoint, and let e_1 and e_2 be two situations with $e_1 < e_2$.
$[[<P_1, P_2>]]^{S, v}(<e_1, e_2>) = 1$ iff $[[P_1]]^{S, v}(e_1) = 1$ and $[[P_2]]^{S, v}(e_2) = 1$,
iff $proj_S(e_1, v) = P_1$ and $proj_S(e_2, v) = P_2$.

This must be generalized to the case in which P_1 and/or P_2 co-occur with musical tunes m_1 and m_2 (which can be treated as trivial in case there is no music), as in (72); for notational simplicity, we write a musical snippet co-occurring with a picture P as a superscript before P, hence: $^{♪m}P$.

(72) Let S be a system of projection and v a viewpoint, and let e_1 and e_2 be two situations with $e_1 < e_2$. Then
$[[<^{♪m1}P_1, {}^{♪m2}P_2>]]^{S, v}(<e_1, e_2>) = 1$ iff $[[^{♪m1}P_1]]^{S, v}(e_1) = 1$ and $[[^{♪m2}P_2]]^{S, v}(e_2) = 1$

The literal content of a picture-music combination is obtained by conjoining the content of the music and of the picture, as in (73).

(73) Let S be a system of projection and v a viewpoint, and let P be a picture co-occurring with music m, whose meaning is $[[m]]^v$, and let e be a situation.[44] Then:
$[[^{♪m}P]]^{S, v}(e) = 1$ iff $[[P]]^{S, v}(e) = 1$ and $[[m]]^v(e) = 1$,
iff $proj_S(e, v) = P$ and $[[m]]^v(e) = 1$

With these truth conditions in hand, we must compute cosuppositional requirements, to the effect that the musical tune co-occurring with a picture should be trivial in the local context of that picture. Let us start with the case in which only the second picture P_2 co-occurs with music, hence the sequence is $<P_1, {}^{♪m2}P_2>$. In different frameworks (such as Heim 1983 and Schlenker 2009), the local context of $^{♪m2}P_2$ is obtained by updating the global context set C with the information contributed by the first picture, to the effect that e_1 satisfies P_1.[45] In this respect, then, we treat the picture sequence in the same way as a conjunction of the form: *$P_1(e_1)$ and $^{♪m2}P_2(e_2)$*, with the second conjunct computed after the first. Usually context sets are treated as sets of possible worlds or possible contexts, but viewpoints, which can be seen as variants of contexts, will be useful: the information contributed by a picture doesn't just pertain to which world we are in, but where in the world we are – hence sets of viewpoints are more appropriate than sets of worlds.

(74) Let C be a context set, treated as a set of viewpoints. Then the local context c' of $^{♪m2}P_2$ in the picture sequence $<P_1, {}^{♪m2}P_2>$ is given by: c' = {v ∈ C: $[[P_1]]^{S, v}(e_1) = 1$} = {v ∈ C: $proj_S(e_1, v) = P_1$}

Next, the cosuppositional requirement in (60) imposes that relative to the local context c' of the second picture P_2, the contribution of the tune co-occurring with P_2 should be trivial, as is stated in (75).

(75) Cosupposition triggered by $<P_1, {}^{♪m2}P_2>$ evaluated in a context set C and two situations e_1 and e_2 with $e_1 < e_2$
For every viewpoint v in c' (as in (74)), if $[[P_2]]^{S, v}(e_2) = 1$, then $[[m2]]^v(e_2) = 1$, i.e. for every viewpoint v ∈ C such that $proj_S(e_1, v) = P_1$, if $proj_S(e_2, v) = P_2$, then e_2 is light-hearted.

Applied to our example, this correctly captures the intuition that if Asterix hits the Roman soldier in e_1, then if he leaves the room in e_2, the latter situation will be light-hearted.

It remains to compute the cosupposition obtained when the music co-occurs with both pictures. We start by computing the cosupposition contributed by the music co-occurring with the first picture. Its local context is just the global context, and so we get the requirement in (76).

---

[44] In accordance with (27), we take musical meaning to be relativized to an auditory point, which we identify with the viewpoint with respect to which the picture is evaluated. This makes non-trivial predictions about correlations between pictorial and musical perspectives, but these are beyond the scope of the present paper and will play no role in our discussions.

[45] We use the terminology of 'update' common in dynamic semantics, but the notion of local context developed in Schlenker 2009 could be employed just as well: it is designed for any system that delivers classical truth conditions and has a well-defined syntax, as is the case for the picture sequences we consider.

(76) Cosupposition triggered by ♯m1P₁ evaluated in a context set C with respect to a situation e₁
 For every viewpoint v in C, if $[[P_1]]^{s,v}(e_1) = 1$, $[[m1]]^v(e_1) = 1$
 iff for every viewpoint v in C, if $proj_s(e_1, v) = P_1$, then e_1 is light-hearted

  Next, we must compute the cosupposition triggered by the music co-occurring with the second picture P_2. Its local context is almost as in (74), except that we must take into account the contribution of the music co-occurring with P_1, as in (77). This gives rise to the cosupposition in (78).

(77) Let C be a context set, viewed as a set of viewpoints. Then the local context c" of P_2 in the picture sequence $<♯m1P_1, ♯m2P_2>$ is given by:
 $c'' = \{v \in C: [[P_1]]^{s,v}(e_1) = 1 \text{ and } [[m1]]^v(e_1) = 1\} = \{v \in C: proj_s(e_1, v) = P_1 \text{ and } [[m1]]^v(e_1) = 1\}$

(78) Cosupposition triggered by ♯m2P₂ in a sequence $<♯m1P_1, ♯m2P_2>$ evaluated in a context set C and two situations e_1 and e_2 with $e_1 < e_2$
 For every viewpoint v in C such that $proj_s(e_1, v) = P_1$ and e_1 is light-hearted, if $proj_s(e_2, v) = P_2$, then e_2 is light-hearted.

 In sum, in case light-hearted music co-occurs with both pictures, we obtain the cosuppositions in (79).

(79) Cosuppositions triggered by $<♯m1P_1, ♯m2P_2>$ evaluated in a context set C relative to a viewpoint v and two situations e_1 and e_2 with $e_1 < e_2$
 (i) For every viewpoint v in C, if $proj_s(e_1, v) = P_1$, e_1 is light-hearted.
 (ii) For every viewpoint v in C such that $proj_s(e_1, v) = P_1$ and e_1 is light-hearted[46], if $proj_s(e_2, v) = P_2$, then e_2 is light-hearted.

  The results are as desired. In our highly simplified analysis, P_1 corresponds to Asterix hitting the Roman soldier and P_2 to Asterix leaving the room. When the lighthearted whistle tune only co-occurs with P_2, we obtain the cosupposition in (75), to the effect that if Asterix hits the Roman soldier in e_1 (= content of the first picture), then if Asterix leaves the room in e_2 (= content of the second picture), e_2 will be light-hearted. By contrast, when the light-hearted tune co-occurs with both pictures, we obtain a stronger cosupposition: if Asterix hits the Roman soldier in e_1, e_1 will be light-hearted; and if Asterix hits the Roman soldier in e_1 (and thus e_1 is light-hearted, by the first cosupposition), then if Asterix leaves the room in e_2, e_2 will be light-hearted.

  We have made radical simplifications for the sake of simplicity, and they should be relaxed in future research: first, we assimilated gif parts to individual pictures; second, we took the content of each picture to pertain to specific situations (e_1 and e_2), whereas their actual content is existential in nature (= there is a certain situation e_1 that projects onto the first picture, and there is a later situation e_2 that projects onto the second picture).[47] Still, the general results of a cosuppositional analysis seem to go in the right direction.

---

[46] The second conjunct is trivially satisfied when the cosupposition in (i) holds.

[47] To be more specific, a more accurate analysis of the truth conditions of the sequence $<P_1, P_2>$ would probably go as in (i), with existential quantification over situations e_1 and e_2:

(i) $[[<P_1, P_2>]]^{s,v} = 1$ iff $\exists e_1 \exists e_2 [e_1 < e_2 \text{ and } proj_s(e_1, v) = P_1 \text{ and } proj_s(e_2, v) = P_2]$

Compositionally, the content of P_1 is plausibly just: $\exists e_1 [proj_s(e_1, v) = P_1]$. By parity of reasoning, the literal content of ♯m1P₁ should be: $\exists e_1 [proj_s(e_1, v) = P_1 \text{ and } m_1(e_1)]$. As a result, a cosupposition is triggered in the scope of an existential quantifier, since we must ensure that, relative to the local context of the nuclear scope (i.e. $[proj_s(e_1, v) = P_1 \text{ and } m_1(e_1)]$), m_1 makes a trivial contribution. Assuming universal projection of presuppositions in existentially quantified structures, as in Heim 1983 and Schlenker 2009, we obtain the condition in (ii):

(ii) ♯m1P₁ triggers a presupposition to the effect that, for each viewpoint v in C, $\forall e_1 [proj_s(e_1, v) = P_1 \Rightarrow m_1(e_1)]$

In our example: every situation in which Asterix hits a Roman soldier (as depicted) is light-hearted.
  We then need to compute the local context of P_2. It will presumably be as in (iii):

(iii) a. Local context of ♯m2P₂ picture in $<P_1, ♯m2P_2>$: $c' = \{v \in C: \exists e_1 proj_s(e_1, v) = P_1\}$
  b. Local context of ♯m2P₂ picture in $<♯m1P_1, ♯m2P_2>$: $c'' = \{v \in C: \exists e_1 [proj_s(e_1, v) = P_1 \text{ and } [[m1]]^v(e_1) = 1]\}$

5.6   *Limitations and possible directions*

There are several limitations to our argument. First, by its very nature it is only an existence proof: it shows that it is possible to find excerpts in which cosuppositional inferences are triggered by the music, not that this is invariably the case. In addition, more systematic empirical investigations should be initiated to (i) assess more precisely the inferences triggered by various excerpts akin to those in (67), and (ii) test their projection behavior in a broader range of environments (e.g. under negation, *might*, *never*, etc.).

Second, even granting that cosuppositions are in fact triggered in our examples, there are at least two ways in which this result could be interpreted. One is that the parasitic nature of music relative to the visuals is responsible for the non-at-issue inference they trigger, just as was argued for co-speech gestures in a different context (Schlenker 2018a, b).[48] Alternatively, it might be that what matters is the semantic content of the music (its emotional character), with the result that because of the implicit "Question under Discussion" (or possibly on even more general grounds, having to do with the projection of emotional inferences[49]) a cosupposition is triggered. We come back to the second possibility in the next section.

Third, the scope of our findings would need to be investigated. One possibility is that it is only to the extent that film excerpts or gifs with music are embedded in a linguistic environment that they trigger cosuppositions. A tantalizing alternative is that the embedding test only serves to reveal a division of information (between at-issue and non-at-issue) that arises even in entirely non-linguistic situations, such as real films or cartoons. One key issue for future research will be to develop presuppositional tests that do not require embedding and can be applied to non-linguistic forms such as films and cartoons.[50]

We then compute the cosupposition triggered by ♪m2P$_2$, focusing for simplicity on the case in (iii)a. The contribution of ♪m2P$_2$ in the context of <P$_1$, ♪m2P$_2$> is presumably of the form in (iv), where the boldfaced part is somewhat problematic: it is needed to establish a connection between the second picture and the first one, but it forces us to adopt a kind of E-type analysis of the relation between situations that satisfy the P$_2$ and P$_1$ (note that as things stand the uniqueness condition of the ι-operator could in principle fail to be satisfied).

(iv)   [∃e$_2$: e$_2$ > **ιe$_1$: proj$_S$(e$_1$, v) = P$_1$**] [proj$_S$(e$_2$, v) = P$_2$ and m$_2$(e$_2$)]

On the assumption that (iv) is correct, the cosupposition triggered by m2P$_2$ is such that, relative to its local context (i.e. within the scope of the existential quantifier over e$_2$), m$_2$ makes a trivial contribution. Assuming once again universal projection of existentially quantified presuppositions, we obtain the result in (v):

(v) In <P$_1$, ♪m2P$_2$>, ♪m2P$_2$ triggers the cosupposition that for every viewpoint v in C such that ∃e$_1$ proj$_S$(e$_1$, v) = P$_1$, [∀e$_2$: e$_2$ > ιe$_1$: proj$_S$(e$_1$, v) = P$_1$] [P$_2$(e$_2$) => m$_2$(e$_2$)].

In words, with some approximations: every situation in which Asterix hits a Roman soldier as depicted is such that any situation that follows it and in which Asterix leaves the room is light-hearted.

[48] Importantly, the expectation that cosuppositional inferences should be triggered by the music only arises to the extent that the music is treated as being parasitic on the visuals and not the other way around. An example in which this is not the case is briefly mentioned in Schlenker, to appear:

(i) On Bastille Day, will your students ♫ *Allons-enfants-de-la-patrie*-HAND-ON-HEART?

In (i), the French words *Allons enfants de la patrie* are literally sung as part of the sentence, but are accompanied by a patriotic posture, with the speaker's hand on his heart. This triggered a cosupposition to the effect that *if the speaker's students were to sing the Marseillaise on Bastille Day, they would adopt a patriotic posture such as having one's hand on one's heart*. In this case, the musical element is primary and the visual (gestural) element is secondary.

[49] This possibility has been emphasized in work by M. Esipova.

[50] The same issues arise for the typology of apparently non-linguistic inferences discussed in Tieu et al.'s 2019, Guerrini and Migotti 2019 and Guerrini and Schlenker 2019.

6   Cosuppositions triggered by music II: pro-speech music

It was recently argued that cosupposition-like inferences are triggered not just when some elements co-occur with and enrich the main message, but also by pure iconic elements such as ASL classifier predicates and pro-speech gestures. While the source of these cosupposition-like inferences is still under investigation, we will argue that similar generalizations can be extended to music.

6.1   Cosuppositions triggered by purely iconic elements

Schlenker, to appear, argues that ASL classifier predicates (whose movement is entirely iconic although the classifier shape isn't) and some English pro-speech gestures trigger cosuppositions despite the fact that they do not involve co-speech or co-sign elements. As an example, (80) involves two realizations of a lifting gesture: first, a neutral lifting gesture, glossed as *LIFT* in (80)a; second a lifting gesture realized with difficulty (trembling hands), glossed as *LIFT-difficult* in (80)b. The paradigm also includes a gesture-free at-issue control, as in (80)c. Acceptability was rated by three native speakers of American English on a 7-pointscale (ratings appear at the beginning of the relevant constructions). The consultants were also asked to assess the strength (also on a 7-point scale) of the cosuppositional inference: *if the speaker were to lift the child, effort/difficulty would be involved*.

(80)   This child, will you

a. [6] *LIFT_*?
Strength of the cosupposition: 1
b. [5.3] *LIFT-difficult*?
Strength of the cosupposition: 4.7
c. [5] lift with difficulty?
Strength of the cosupposition: 1.3
(video 01, a,b,d; 3 consultants; from Schlenker, 2018c and Schlenker, to appear)

The results suggest are indicative a weak cosupposition under questions with *LIFT-difficult* in (80)b but not with the at-issue control in (80)c. Embedding under other operators confirms this pattern of projection (Schlenker 2018c, Schlenker, to appear).

We submit that the same effects arise with entirely different iconic forms. If we minimally modify minimally our example from (56) so that the pictorial element now becomes a pro-speech rather than an co-speech image, as in (81) we obtain a fairly clear meaning: the question is whether Asterix will drink the magic potion.[51]

(81)

(There is an impending battle, but Asterix has magic potion with him.)

What will happen next: will Asterix …

But something else is inferred as well. Even for a reader who is not familiar with Asterix (we think), there is likely to be an inference to the effect that *if Asterix drinks the magic potion, he will do so with the effects depicted*. Intuitively, what is going on is that the picture provides way too many details for the question to be whether Asterix will drink the magic potion *in this precise way*. Rather, the question

---

[51] The picture (which is not from Asterix's creator Uderzo) can be found at https://www.deviantart.com/zenitram-anth/art/Asterix-chez-les-freaks-472781613 (December 9, 2019).

is whether Asterix will drink the magic potion, and the assumption conveyed is that *if he does so, the effects will be as depicted*.

 While several theories could be considered to explain why purely iconic elements trigger presuppositions, a partly non-unified theory of cosuppositions was proposed in Schlenker, to appear. The idea, stated in (82), is that there are two somewhat different reasons why an entailment might be presented as being "unimportant", an undefined notion in (60) above. First, an entailment could be presented as unimportant because it is contributed by a secondary message, one that is parasitic on the main message: this is the case of co-speech and co-sign gestures, and arguably of co-film music. Second, however, an entailment could be presented as unimportant for conceptual reasons, for instance because it fails to answer the question under discussion.

(82) An entailment p' might be presented as unimportant for different reasons:
(i) for reasons of manner, in case p' is contributed by a co-speech or co-sign gesture (which is parasitic and thus should not make an essential contribution);
(ii) for conceptual reasons, in case p' is understood not to matter given the context of the conversation.
(Schlenker, to appear)

 To illustrate, if the implicit Question under Discussion is whether Asterix will drink the magic potion, we are faced with the situation depicted in (83): the question introduces two cells, corresponding to the possible worlds in which Asterix drinks the magic potion (on the left), and those in which he doesn't (on the right). But due to its iconic content, the picture doesn't just provide information about the fact that Asterix will drink the magic potion, but also about *how* he will do so. As a result, if we go by the literal meaning of the picture, while a 'yes' answer would settle the question, a 'no' answer wouldn't: it could be that Asterix won't drink the potion (right-most cell), or that he will drink it, but not as depicted (top-most on the left).

(83) **Will Asterix drink the magic potion?**

Asterix will drink the magic potion   Asterix will not drink the magic potion

The generation of a cosupposition is taken to be the minimal pragmatic way to address the problem: by positing a presupposition that *if Asterix drinks the magic potion, he will do so as depicted*, as shown in (84).

(84) Writing Asterix will drink the magic potion as p and Asterix will drink the magic potion as depicted as pp', in order to guarantee that the question under discussion is addressed by a 'no' (as well as a 'yes') answer in (83), we need a presupposition that we are not in the hatched area, correspoding to $(p \wedge \neg pp')$, hence: $\neg(p \wedge \neg pp')$, which simplifies to $(\neg p \vee pp')$, i.e. $p \Rightarrow p'$.

 If this analysis is on the right track, we expect that cosuppositions could, given the appropriate context, be triggered by pro-speech music as well, i.e. by music that replaces some words. We turn to initial data that suggest that this might be on the right track.

6.2 *Cosuppositional effects with pro-speech music?*

An initial example of a possible cosuppositional effect triggered by pro-speech music is displayed in (85)b. It features the first phrase of Beethoven's Für Elise (a staple of the piano student repertory), but intentionally played badly. It arguably triggers the inference that if the student plays this piece, he will

do so roughly as illustrated, i.e. badly. This contrasts with a pro-speech musical excerpt featuring a standard rendition of the same notes, in (85)a, and also with an at-issue control in which the bad interpretation is used with the expression *like this*, as in (86)b.

(85) Which piece will your new student play this afternoon? Will he
    a. ELISE ?   https://www.dropbox.com/s/mkaahi3v8mrc50o/Beethoven-Elise-normal.m4a?dl=0
    b. ELISE'?   https://www.dropbox.com/s/ud2kdw4ut2wv6tv/Beethoven-Elise-bad.m4a?dl=0
    b => if he plays, he will do so as shown

(86) Will your new student play like this?
    a. ELISE ?   https://www.dropbox.com/s/mkaahi3v8mrc50o/Beethoven-Elise-normal.m4a?dl=0
    b. ELISE'?   https://www.dropbox.com/s/ud2kdw4ut2wv6tv/Beethoven-Elise-bad.m4a?dl=0
    b ≠> if he plays, he will do so as shown

The inference that is arguably derived in (85)b is unsurprising in view of the analysis of pro-speech cosuppositions sketched above: the Question under Discussion is which piece the student will play. The literal meaning of the verbal component in (85)b, realized by a musical excerpt, is overly specific, as it contributes the information that the student will play Für Elise, and that he will play it badly. This is the same problem we saw in (83), and it is solved by the same pragmatic means, namely by the generation of a presupposition that if the student plays, he will do so as shown.

This initial example is somewhat limited, however, because it is quasi-quotational in nature: the excerpt denotes an action of playing that very excerpt. But non-quotational cases can arguably be constructed as well. We start from (87)a, where the beginning of Rossini's William Tell Overture is used to evoke the arrival of the cavalry; due to the fast rendition of the music, this might already trigger a cosupposition that if the cavalry comes riding in, they will do so skillfully. By contrast, in (87)b the excerpt is played with numerous wrong notes, and this arguably triggers the cosupposition that if the cavalry comes riding in, they'll do so in an unskilled fashion. Finally, (87)c features a slow rendition, which becomes even slower as the excerpt progresses, which suggests that if the cavalry comes riding in, they'll do in a slow fashion, and probably with difficulty.

(87) [Phlegmatic pianist, to the mayor of a besieged city]
    Sir, I am told the enemy is about to enter the city. Will we be saved? Will our old cavalry…
    a. TELL-fast ?   https://www.dropbox.com/s/vbsksds7y0wnqqb/Rossini-Tell%20original-louder.mp4?dl=0
    => if our old cavalry comes riding in, they'll do so skillfully
    b. TELL-wrong_notes ?   https://www.dropbox.com/s/p3ic6p8yq2rf3ow/Rossini-Tell%20avec%20dissonances-80-louder.mp4?dl=0
    => if our old cavalry comes riding in, they'll do so in an unskilled fashion
    c. TELL-slowing_down ?   https://www.dropbox.com/s/3rogxqnivub86gn/Rossini-Tell%20ralentissement3-louder.mp4?dl=0
    (or: TELL-slowing_down-alternative) https://www.dropbox.com/s/jzmour4p9pnm9xa/Rossini-Tell%20ralentissement2-louder.mp4?dl=0
    => if our old cavalry comes riding in, they'll do so in a slow fashion

These examples would need to be assessed with experimental means in the future, something that hasn't been done yet even for cosuppositions generated by pro-speech gestures. Pending further investigation, we conclude that pro-speech music might be able to trigger cosuppositions in the same pragmatically determined cases as pro-speech gestures (as well as ASL classifier predicates) and possibly even drawings.[52]

The same disclaimers apply as in our discussion of co-film music. Even if our conclusion is on the right track, it only shows that musical excerpts can trigger cosuppositions *when they are embedded in sentences*. This is compatible with the stronger claim that even without a linguistic environment similar cosuppositions can be triggered by pure music, but this conclusion does not follow from our data, which exclusively pertain to pro-speech music.

---

[52] Needless to say, this finding only highlights the ambiguity of findings on co-film and co-gif music discussed in the preceding section: these too could in principle be due to the implicit Question under Discussion rather than to the parasitic character of the music on the visuals.

7 Conclusion

7.1 *Main results*

In sum, we have restated and hopefully clarified initial claims about the existence of a music semantics, and we have proposed that it can be enriched along two dimensions by borrowing ideas from pictorial semantics and from gesture semantics.

Our first claim was that the existence of a music semantics is not at all threatened, but in fact clarified, by the observation that even program music cannot tell stories with anything like the level of specificity it purports to have. Bernstein famously inferred from his ability to tell a story about Superman for a piece intended to evoke Don Quixote (in Strauss's Variation II) that the true meaning of music is "the way it makes you feel when you hear it" . Nothing of the sort follows. Bernstein's own Superman story was mostly isomorphic to the Don Quixote story intended by Strauss. This is no accident: the details of the music conspire to trigger definite if abstract inferences, as we illustrated by studying the role of melodic movement, dissonances, loudness, and a final cadence. The source of these inferences can be brought out by using the method of minimal pairs: by recomposing the music so as to modify one parameter at a time while remaining faithful to the rules of the genre, we were able to display the source of several important inferential effects.,

Our second claim was that the initial 'toy model' of music semantics offered in Schlenker 2017, 2019a is insufficient. Migotti 2019 correctly observed that defining the semantics in terms of preservation of certain orderings among some musical properties (such as loudness, frequency, etc) is too weak. But in addition, a crucial idea can be borrowed from the semantics of visual narratives: there are crucial ambiguities in music as in visual narratives pertaining to relations of cross-reference among objects. The surface of the music can help make some patterns of cross-reference relations more or less plausible. But decisions about these seem to be crucial in music performance, orchestration, and setting of dance to music.

Our third claim pertained to cases in which music accompanies another medium which can be taken to be primary in the transmission of a message. This is another incarnation of a situation that has been investigated in some detail in research on co-speech or co-sign gestures and facial expressions. Recent literature has argued that these gestural expressions trigger cosuppositions, and speculated that this is because they are presented as parasitic on the message that they enrich (both parts are still the topic of active debates, e.g. Ebert and Ebert 2014 and Esipova 2019). Pasternak 2019 extended these findings to co-speech sounds, and a Pasternakian extension to co-speech music is immediate. Moving outside of language, we suggested that co-film music might trigger cosuppositions as well. In order to make this point, we investigated composite utterances made of words combined with pro-speech film excerpts or gifs, which could then be combined with different kinds of music. While the specific semantic enrichments depended on the music chosen, it seemed clear that, in the cases we considered, the contribution of the music was not at-issue, and was better analyzed as being cosuppositional in nature. In this case, the cosuppositional character of the inference might be due to the parasitic character of the music (although alternative theories are possible as well). Still, there are further cases in which cosuppositions are triggered by pro-speech music, which by definition couldn't be parasitic on anything (because it fully replaces a word). These cases are conceptually and empirically similar to cosuppositions triggered (in restricted pragmatic conditions) by some pro-speech gestures and possibly even drawings.

One key question for future research is whether cosuppositional effects exist in film or cartoon music that is not embedded in a linguistic environments.

7.2 *Broader conclusions*

Two points can be made from a broader perspective. First, initial formal attempts emphasized how different music semantics is from linguistic semantics; the semantics of visual animations was presented as a far better point of comparison (Schlenker 2017, 2019a). Still, the existence of discourse referents in music (as in visual narratives) makes its semantics a tiny bit more linguistic-like than was initially surmised. The cosuppositional nature of co-film and co-gif music further highlights a similarity with some phenomena that are found in language. But in both cases, the relevant natural class is almost

certainly broader than just language and music (in fact, it was existence of discourse referents in visual narratives that led to the exploration of similar issues in music semantics).

Second, we hope that our explorations might highlight the fruitfulness of the generalized approach to meaning associated with Super Semantics. The initial motivation for this endeavor was in part methodological, as the investigation of meaning as truth conditions can naturally be extended to several non-traditional representational forms. But this generalized approach also makes it possible to draw unexpected connections among very different semantic systems. The analysis of musical meaning has thus been enriched by the investigation of discourse referents in visual narratives and by gesture theory; conversely, theories of cross-reference and of cosuppositions are made empirically richer thanks to music semantics.

References

Abusch, Dorit: 2013, Applying discourse semantics and pragmatics to co-reference in picture sequences. In *Proceedings of Sinn und Bedeutung* 17: 19–25.

Abusch, Dorit: 2015, Possible Worlds Semantics for Pictures. Manuscript, Cornell University.

Abusch, Dorit and Rooth, Mats: 2017, The formal semantics of free perception in pictorial narratives. In Alexandre Cremers, Thom van Gessel & Floris Roelofsen (eds), *Proceedings of the 21st Amsterdam Colloquium*. Available online at https://semanticsarchive.net/Archive/jZiM2FhZ/AC2017-Proceedings.pdf.

Bedoya, Daniel: 2019, Les émotions sont-elles exprimées de la même façon en musique que dans la voix parlée? Internship report (advisors: Jean-Julien Aucouturier and Louise Goupil), IRCAM, Paris.

Bernstein, Leonard: 1967, Charles Ives: American Pioneer. Young People's Concerts. Television series, February 23, 1967.

Bernstein, Leonard: 1976, *The Unanswered Question: Six Talks at Harvard*. Harvard University Press.

Bernstein, Leonard: 2005, *Young People's Concerts*. Foreword by Michael Tilson Thomas. Amadeus Press, Pompton Plains, New Jersey.

Bregman, Albert S.: 1994, *Auditory Scene Analysis*. MIT Press.

Cooper, Michael: 2013, Mystery of the Missing Music. New York Times, August 27, 2013. Retrieved online at https://www.nytimes.com/2013/08/28/arts/music/benjamin-brittens-lost-score-for-les-sylphides.html on December 18, 2019.

Craine, Debra and Mackrell, Judith: 2010, *Oxford Dictionary of Dance*. Oxford University Press.

Ebert, Cornelia and Ebert, Christian: 2014, Gestures, Demonstratives, and the Attributive/Referential Distinction. Handout of a talk given at Semantics and Philosophy in Europe (SPE 7), Berlin, June 28, 2014.

Ebert, Cornelia: 2018, A comparison of sign language with speech plus gesture. *Theoretical Linguistics* 44(3-4):239-249.

Esipova, Maria: 2019, *Composition and projection in speech and gesture*. PhD thesis, New York University.

Granroth-Wilding, Mark and Steedman, Mark: 2014, A robust parser-interpreter for jazz chord sequences. *Journal of New Music Research* 43 (4), 355-374.

Greenberg, Gabriel: 2013. Beyond Resemblance. *Philosophical Review* 122:2, 2013

Greenberg, Gabriel: 2014, Reference and Predication in Pictorial Representation. Handout of a talk given at the London Aesthetics Forum (February 19, 2014).

Greenberg, Gabriel: 2019, Tagging: Semantics at the Iconic/Symbolic Interface. *Proceedings of the Amsterdam Colloquium 2019*.

Guerrini, Janek and Migotti, Léo: 2019, Musical gestures in the typology of linguistic inferences. Talk given at the workshop on "Linguistic investigations beyond language: gestures, body movement and primate linguistics", March 3, 2019.

Guerrini, Janek and Schlenker, Philippe: 2019, Linguistic inferences without words: the case for pro-speech vocal gestures. Poster, GLOW 42, May 8-10, 2019.

Hanslick, Eduard: 1891, *The Beautiful in Music: a Contribution to the Revisal of Music Aesthetics*. Translated by Gusav Cohen. London: Novello and Company, Limited.

Heider, F., and Simmel, M.: 1944, An experimental study of apparent behavior. *American Journal of Psychology*, 57, 243-259.

Heim, Irene: 1983, On the projection problem for presuppositions. In Barlow, M. and Flickinger, D. and Westcoat, M. (eds.), *Second Annual West Coast Conference on Formal Linguistics*, pp. 114–126, Stanford University.

Huron, David: 2016, *Voice Leading: The Science Behind a Musical Art*. MIT Press.

Keats, Jonathan: 2018, Science of Music: Listen up! *Discover Magazine*, special issue on *Everything worth knowing,* August 2018. http://discovermagazine.com/2018/jul-aug/science-of-music

Koelsch, S.: 2012, 'Musical Semantics'. Chapter from *Brain and Music*, Wiley-Blackwell.

Kominsky, J.F., Strickland, B., Wertz, A.E., Elsner, C., Wynn, K., & Keil, F.C.: 2017, Categories and constraints in causal perception. *Psychological Science* 28,11: 1649-1662

Lascarides, Alex and Stone, Matthew: 2009. A Formal Semantic Analysis of Gesture. *Journal of Semantics*, 26(4):393-449.

Lerdahl, Fred and Ray Jackendoff: 1983, *A generative theory of tonal music*. Cambridge, MA: MIT Press.

Lerdahl, Fred: 2001. *Tonal Pitch Space*. Oxford University Press.

Lerdahl, Fred: 2019, *Composition and Cognition: Reflections on Contemporary Music and the Musical Mind*. University of California Press.

Mehr, SA et al.: 2019, Universality and diversity in human song. *Science* 366, eaax0868:1-17, 2019.

Meyer, L.B.: 1956, *Emotion and Meaning in Music*. University of Chicago Press, Chicago.

Migotti, Léo: 2019, *Towards a theory of music semantics*. MA thesis, Cogmaster, Paris.

Migotti, Léo and Zaradzki, Léo: 2019, Walk-denoting music: refining music semantics. To appear in the *Proceedings of the Amsterdam Colloquium 2019*.

Nudds, Matthew: 2007, Auditory Perception and Sounds. Manuscript.

Pasternak, Robert: 2019, The Projection of Co-speech Sound Effects. Manuscript, ZAS Berlin. https://ling.auf.net/lingbuzz/004520

Pesetsky, David and Katz, Jonah. 2009. The Identity Thesis for Music and Language. Manuscript, MIT.

Rohrmeier, Martin: 2011, Towards a generative syntax of tonal harmony. *Journal of Mathematics and Music* 5 (1), 35–53.

Schlenker, Philippe: 2009, Local Contexts. *Semantics & Pragmatics,* Volume 2, Article 3: 1-78, doi: 10.3765/sp.2.3

Schlenker, Philippe: 2010, Presuppositions and Local Contexts. *Mind* 119, Issue 474: 377-391

Schlenker, Philippe: 2011, Presupposition Projection: Two Theories of Local Contexts – Parts I and II. *Language and Linguistics Compass* 5, 12: 848–857 and 858–879.

Schlenker, Philippe: 2017, Outline of Music Semantics. *Music Perception: An Interdisciplinary Journal* 35, 1: 3-37 DOI: 10.1525/mp.2017.35.1.

Schlenker, Philippe: 2018a, Gesture Projection and Cosuppositions. *Linguistics & Philosophy* 41, 3:295–365.

Schlenker, Philippe: 2018b, Iconic Pragmatics. *Natural Language & Linguistic Theory* 36, 3:877–936

Schlenker, Philippe: 2018c, Sign Language Semantics: Problems and Prospects [replies to peer commentaries]. *Theoretical Linguistics* 44(3-4): 295–353.

Schlenker, Philippe: 2019a, Prolegomena to Music Semantics. *Review of Philosophy & Psychology* 10 (1): 35-111. https://doi.org/10.1007/s13164-018-0384-5

Schlenker, Philippe: 2019b, What is Super Semantics? *Philosophical Perspectives* 32, 1: 365-453 https://doi.org/10.1111/phpe.12122

Schlenker, Philippe: 2019c, Gestural Semantics: Replicating the typology of linguistic inferences with pro- and post-speech gestures. *Natural Language & Linguistic Theory* 37, 2: 735–784.

Schlenker, Philippe: to appear, Iconic Presuppositions. *Natural Language & Linguistic Theory*.

Shawn, Allen: 2014, *Leonard Bernstein: an American Musician*. Yale University Press.

Tieu, Lyn; Pasternak, Robert; Schlenker, Philippe; Chemla, Emmanuel: 2017, Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa* 2(1).

Tieu, Lyn; Pasternak, Robert; Schlenker, Philippe; Chemla, Emmanuel: 2018a, Co-speech gesture projection: Evidence from inferential judgments. *Glossa* 3(1), 109. DOI: http://doi.org/10.5334/gjgl.580

Tieu, Lyn; Schlenker, Philippe; Chemla, Emmanuel: 2019, Linguistic Inferences Without Words. *Proceedings of the National Academy of Sciences* 116 (20) 9796-9801