Musical Meaning within Super Semantics*

Philippe Schlenker**

March 16, 2021

To appear in *Linguistics & Philosophy*

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 three 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 we highlight it by investigating ways in which orchestration and dance may make cross-referential dependencies more explicit. Second, we show that by bringing music semantics closer to the semantics of visual narratives, we can give an account of the semantics of mixed visual and musical sequences. Third, 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

* **Music consultant: Arthur Bonetto.** Arthur Bonetto served as a regular and very insightful music consultant for these investigations; virtually all music examples were discussed with him, and he played a key role in the construction of all minimal pairs, especially when a piece had to be rewritten with special harmonic constraints. However he bears no responsibility for the theoretical claims – and possible errors – contained in this piece.

Acknowledgments: I am very grateful to Paul Egré for initial conversations on Bernstein's views on musical meaning (and on his Superman example), and to Emmanuel Chemla and Pritty Patel-Grosz for extremely helpful written comments on the manuscript. For very helpful feedback, I am also grateful to audiences at GLOW 2019 (May 11, 2019) and NYU's CLaME group (October 15, 2019). I received very helpful suggestions from Amir Anvari, Paul Boghossian, Emmanuel Chemla, Jeremy Kuhn, Léo Migotti, Robert Pasternak, Pritty Patel-Grosz, Lyn Tieu, and of course Arthur Bonetto. I am also grateful to Zenitram for giving me authorization to use his picture of Asterix. Léo Migotti did considerable (and greatly appreciated) work to make audio-visual examples maximally available in the final version. Lastly, this paper was greatly improved following extraordinarily constructive comments and criticisms from two anonymous reviewers for *Linguistics & Philosophy*, and from Managing Editor Regine Eckardt.

Grant acknowledgments: This research received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 788077, Orisem, PI: Schlenker). Research was conducted at Institut d'Etudes Cognitives, Ecole Normale Supérieure - PSL Research University. Institut d'Etudes Cognitives is supported by grant FrontCog ANR-17-EURE-0017.

** Institut Jean-Nicod - CNRS, UMR 8129, ENS/EHESS - PSL Research University F-75005 Paris, France. Department of Linguistics, New York University, New York.

1 Introduction 3

2 Music semantics: a source-based analysis 4

2.1 Main ideas 4

2.2 Schematic illustration 6

2.3 Testing semantic effects 8

2.4 Objections 9

3 An objection to music semantics: Bernstein's challenge 10

3.1 'No semantics' 10

3.2 Bernstein's objection 11

3.3 Music semantics in action: minimal pairs 13

3.4 How abstract is musical meaning? 17

4 Discourse referents: from pictorial semantics to music semantics 17

4.1 Music semantics vs. picture semantics 18

4.2 Adding discourse referents to pictorial semantics 19

4.3 Adding discourse referents to music semantics 21

4.4 Revealing discourse referents: an example 23

4.5 Creating minimal pairs 27

4.6 Syntactic vs. semantic analyses 28

4.7 Intermediate conclusion and refinements 29

5 Music combined with visual animations 30

5.1 An example from Disney's Sorcerer's Apprentice 30

5.2 Truth of mixed sequences 31

5.3 Application to the example 31

5.4 Derived notions 32

5.5 Further questions 32

6 Cosuppositions triggered by music I: co-speech and co-film/gif music 32

6.1 Gestural cosuppositions 33

6.2 The generality of cosuppositions 34

6.3 Co-film music can trigger cosuppositions 36

6.4 Refining the argument 38

6.5 Analyzing the contrasts 40

6.6 Limitations and extensions 42

7 Cosuppositions triggered by music II: pro-speech music 42

7.1 Cosuppositions triggered by purely iconic elements 42

7.2 Cosuppositional effects with pro-speech music? 45

7.3 Can conceptual considerations fully explain cosuppositions triggered by music? 46

8 Conclusion 47

8.1 Main results 47

8.2 Broader conclusions 48

Appendix. Rule and Derivations 49

References 61

1 Introduction[1]

The study of musical syntax with formal means has by now a well-established tradition (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, Zaradzki 2021). This is part of a more general attempt to apply the general methods of formal semantics beyond human language (sometimes called 'Super Semantics'[3], see for instance Greenberg 2013, Abusch 2013, 2019, Schlenker 2019b). First, there are systematic ways in which music triggers inferences about a music-external reality: music semantics has an object. Second, music can convey information about objects undergoing events thanks to a more abstract version of inferential mechanisms that produce information about sound sources in general. More precisely, 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 an object evoked (or 'denoted') by the music is losing energy or moving away from the perspectival center. Other inferences are triggered by specifically musical properties, as when a dissonance is taken to signal that the denoted object is in 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.

The present piece has four goals. After summarizing for a linguistics audience the main claims of music semantics (Section 2), we address a common objection according to which music has no meaning besides the emotions it triggers in the listener (Section 3). Specifically, we consider 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 three 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.[4] We argue that a similar conclusion extends to music, and we highlight it by investigating ways in which orchestration and dance can make patterns of cross-reference more explicit (Section 4). Second, we make use of these tools (including discourse referents) to propose an initial semantics for mixed sequences, such as pictorial animations combined with music (Section 5). Third, it has been 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

[1] Audiovisual examples have been included by way of URLs (some are borrowed from Schlenker 2017, 2019a,b). They can also be downloaded as a separate file, and are cross-referenced in the text by way of boldfaced names such as **AV00**, **AV01**, …. The text might be hard to follow without consulting these audiovisual examples.
Downloadable file: https://drive.google.com/file/d/1k4-6296WVOP32LZtBuiihzXgK4S3xHxS

[2] This notion of semantics corresponds to what Koelsch 2012 calls 'extra-musical meaning'. Granroth-Wilding and Steedman 2014 endow their formal syntax for jazz chord sequences with a semantics that encodes paths in a tonal pitch space; this does not yield an 'extra-musical meaning' in Koelsch's sense, or in the sense of 'semantics' adopted in this piece (see also Schlenker 2019, Appendix I).

[3] Gabe Greenberg has used the term *formal semiotics* with a related meaning (e.g. Greenberg 2021, and numerous earlier talks).

[4] Our 'discourse referents' will just be variables with a standard semantics; the framework we develop does not make use of notions borrowed from dynamic semantics (unlike Abusch 2019) or Discourse Representation Theory (unlike Maier and Bimpikou 2019). See also Greenberg 2014, 2019b for a different way of introducing some varieties of discourse referents in pictorial representations.

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. very brief silent animations[5]) in a sentence in order to test presupposition projection (Section 6). Relatedly, 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 suggest that the same conclusion might apply to some examples of pro-speech music (Section 7). For reasons of readability, we have taken some shortcuts at several junctures, but an Appendix gives a more explicit version of the rules and especially semantic derivations we are argue for in this piece.

Taken together, our results will 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.

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, Migotti and Zaradzki 2019, and Zaradzki 2021. For the sake of clarity and brevity, we restrict attention to a "bare-bones" music semantics.[6]

Throughout, music semantics takes as input the realization of a piece rather than just written notes, which are underspecified in numerous respects. In the same way, linguistic semantics takes as input articulated rather than just written words (which lack intonation and numerous semantically relevant properties): in musical as well in linguistic analyses of meaning, the written form has no special theoretical status.

2.1 Main ideas

Recent music semantics treats music as a kind of abstract auditory animation. In auditory or visual perception, the subject seeks 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 hear a low-frequency call produced by an animal and infer that it is very large; or one may hear a car engine whose loudness decreases and infer that it is moving away. The same general idea applies to music semantics: the listener seeks to draw inferences about certain objects, which share some properties with the sources of the musical sounds.[7] But music semantics is special in several respects (see Schlenker 2019a for a more detailed discussion).

First, only some of the properties of the sound are interpreted. If all were, when we hear a double bass we would just infer… that we are hearing a double bass. But its low-frequency sound is often used to evoke large objects, as in Saint-Saëns's Carnival of Animals, where it represents an elephant. Similarly, instruments playing diminuendo (with diminishing loudness) can be used to evoke something moving away. If all properties of the sound were interpreted, we would just infer that the musicians are playing more softly, not that anything is moving away. In music semantics, then, one uses some but not all properties of the music to draw inferences about objects and what happens to them.

It may be helpful to compare musical animations to abstract visual representations. A very simple example of a static one (from Schlenker 2017, 2019a) appears in (1): three columns of various

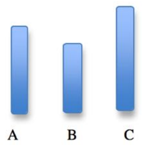
[5] *Gif* just stands for *Graphic Interchange Format*, but the term has come to be used to refer to very brief animations.

[6] More sophisticated notions pertaining to the interface between music syntax and semantics, or semantics and pragmatics, are discussed at varying levels of detail in Schlenker 2017, 2019a, Zaradzki 2021.

[7] The term "virtual source" has been used by different authors to discuss inferential effects in music. Thus for Bregman 1994, "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). We eschew the term here because it has been used in different ways by different authors, leading to potential confusions (e.g. Schlenker 2017, 2019 uses the term to refer to the objects – not necessarily sound-producing ones – which are denoted by the music).

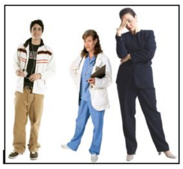
heights (A, B, C), arranged from left to right, are used to depict individuals as in the scenes in (2), involving a boy, a doctor and a business woman.

(1) A pictorial representation

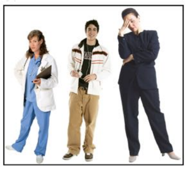


(2) Three possible denotations for (1)

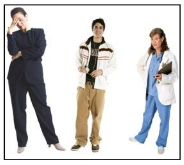
a.



b.



c.



The idea was that the columns can more readily represent the scene in (2)a than those in (2)b,c because in the former case but not in the latter both the ordering by height of the columns and their left-to-right order is preserved (using the denotation relations: A → the boy, B → the doctor, B → the business woman). It is clear that if all visual properties of the columns were interpreted, we would just conclude that they can represent none of the scenes in (2), since the columns don't look like human beings.

Second, just like our columns could stand for very diverse objects, so too do musical animations have an extremely underspecified semantics. As we will see in Section 3, even 'program music', written with the explicit goal of evoking a particular story, has a meaning that is far more general and can thus be made to fit entirely different stories (although ones that share some structural properties with the original one). This is the sense in which musical meaning is abstract: it is true of a large set of diverse situations, just like our three columns in (4) could be taken to evoke very diverse objects.[8]

Third, and relatedly, inferences may be about objects that are not sound-producing. This abstractness is essential because 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).[9] This result is achieved because the inferential mechanisms at play allow sound properties to produce information about objects that are not implicated in sound production. For instance, decreasing loudness may be interpreted in terms of an object losing energy or moving away, and both events may characterize entities that are silent (this point will become more vivid when we illustrate the workings of the system in Section 2.2).

Fourth, inferential rules are of two kinds. Some are lifted from normal auditory cognition, as in the case of a low-frequency sound used to evoke a large object. 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.[10]

These inferential rules can be illustrated as follows (where 'virtual source' refers to the denoted object):

[8] In our technical implementation, 'situations' are cashed out in terms of possible worlds and tuples of eventualities.

[9] Similarly, the beginning of Richard Strauss's Thus spoke Zarathustra, discussed in Schlenker 2017, 2019a, was intended to evoke a sunrise.

[10] We do not seek to do justice to the complexity of harmonic notions, and in this respect our discussion is *greatly* simplified. See for instance Lerdahl 2001 for a detailed discussion of tonal pitch space (and Steedman 2002 for some historical remarks).

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.[11] 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 cases are discussed in Schlenker 2017, 2019a.[12] We provides examples in (3) to show that musical inferences have real substance, but also to highlight that their general form is that in (4), which pertains to the preservation of some relations among musical events in the space of denoted (i.e. world) events.

(3) Examples of inferential effects (originally from Schlenker 2019b Appendix II, with links to examples)[13]
a. Lower pitch may indicate that the denoted object (i) is larger, or (ii) is less excited/energetic.
b. Lower loudness may indicate that the denoted object is (i) less energetic, or (ii) further away.
c. Lower speed may indicate that the denoted object is slower.
d. Silence may indicate that an event is interrupted.
e. Lesser harmonic stability may indicate that the denoted object is in a less stable (i) physical or (ii) emotional position.

(4) If musical events M_1 and M_2 stand in relation R (i.e. M_1 R M_2), their respective denotations e_1 and e_2 stand in relation R* (i.e. e_1 R* e_2).

We may think of this observation in the following terms: a musical voice can denote an object undergoing a series of world events, one for each of the musical events. But not anything goes: the world events must preserve certain orderings that are found among the musical events. To illustrate, (3)b starts from an ordering of musical sounds by loudness, and states that if musical event M_1 is less loud than a musical event M_2, the respective world events they denote, e_1 and e_2, should either guarantee that the denoted object (i) is less energetic in e_1 than in e_2, or (ii) is further from a certain perspectival point in e_1 than in e_2 (the role of the world and of the perspectival point will be made more explicit below).

There might well be stronger semantic requirements. For instance, even a single low-pitch sound may be indicative of a large object, and preservation conditions as in (4) will have no 'bite' for a single sound (there should be at least two for them to have an effect). So an analysis in terms of preservation conditions does not claim to be exhaustive, but just to offer a 'proof of concept' of what a music semantics could be.

2.2 Schematic illustration

We will now make these ideas more precise with a slightly modified version of rules and examples posited in Schlenker 2019b (a more explicit version can be found in the Appendix, which revisits these ideas within the system with variables developed in later sections). We restrict attention to the case in which a musical sequence is taken to provide information about a single object and the events it

[11] 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.

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

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

undergoes. We use the cover term 'eventualities' to refer to events and to states (e.g. Parsons 1990); events are often more natural when discussing the semantics of music (as sounds are typically interpreted as indicative of something that happens), while states are more intuitive when talking about the semantics of pictures, as we will see later in this piece.

We start from a sequence $<M_1, \dots, M_n>$ of n (adjacent) musical events, ordered in time as they are in the sequence. The first step is to state that a sequence of n musical events is true of an object O and n eventualities if (i) O takes part in all n eventualities, (ii) these are ordered in time in the same way as the musical events, and (iii) O and the eventualities obey certain preservation conditions. For simplicity, we consider just two: loudness, whose interpretation is lifted from normal auditory cognition, and harmonic stability, which is more purely musical in nature (throughout, *iff* abbreviates *if and only if*):

(5) **Truth-of relative to a perspectival point and a world**
Let π be a perspectival point, and w a world. Then:
A musical sequence $<M_1, \dots, M_n>$ is true of an object O and of eventualities $<e_1, \dots, e_n>$ relative to π in w iff in w, for each k such that $1 \leq k \leq n$, O takes part in e_k, and
(1) temporally, $e_1 < \dots < e_n$;
(2) relative to π, and w, the Loudness and Harmonic stability conditions are satisfied by O and $<e_1, \dots, e_n>$ with respect to $<M_1, \dots, M_n>$.

Truth-of is relativized to a world, as is standard in model-theoretic semantics, and also to a perspectival point. The latter matters because some conditions make reference to the relation between the denoted object O and a perceiver; for instance, decreasing loudness may be interpreted as an object moving away from the perceiver. Perspectival points will thus have to include information about spatial points (for pictorial applications, we will develop a more fine-grained notion, but it will still include information about a spatial point and will thus be appropriate for musical applications as well.[14])

The next step is to state our preservation conditions. Here our simple-minded semantics solely seeks to preserve certain orderings, as suggested above by (3)-(4). For very abstract representations, such as the diagrammatic columns in (1) (preservation of ordering by height and of left-to-right order), this may be accurate. For music, this will turn out to be too simple, but this statement has the advantage of being formally manageable.

(6) **Preservation conditions**
Relative to a perspectival point π and a world w, if for each $i \leq n$ the object O takes part in eventuality e_i, a musical sequence $<M_1, \dots, M_n>$ is true of O and of eventualities $<e_1, \dots, e_n>$ only if O and $<e_1, \dots, e_n>$ satisfy the following preservation conditions with respect to $<M_1, \dots, M_n>$.
a. Loudness condition
For all $i, k \leq n$, if M_i is less loud than M_k, then in w either:
(i) O has less apparent energy from the perspective of π in e_i than in e_k; or
(ii) O is further from π in e_i than in e_k.

b. Harmonic stability condition
For all $i, k \leq n$, if M_i is less harmonically stable than M_k, then from the perspective of π in w O is in a less stable position in e_i than in e_k.

(We have stated condition (6)a(ii) in terms of *apparent* energy from a certain perspective because we are interested in a psychological notion; we will for instance state that a sun rising appears to gain energy, irrespective of what the underlying physics says.)

From (5)-(6), we may derive a definition of truth in a world relative to a perspectival point by existentially quantifying over objects and tuples of eventualities, as in (7).

(7) **Truth relative to a perspectival point and a world**
Let π be a perspectival point and w a world. Then:
A musical sequence $<M_1, \dots, M_n>$ is true relative to π in w iff for some object O and for some eventualities $e_1, \dots, e_n$, $<M_1, \dots, M_n>$ is true of O and $<e_1, \dots, e_n>$ relative to π and g in w.

Since truth (as in (7)) is easy to derive from truth-of (as in (5)), but more cumbersome, we will usually be content to derive truth-of conditions in our discussions.

[14] To avoid having to redefine later what perspectival points are, we do not *identify* them with spatio-temporal points; at this point we just require that they include information about a spatial-temporal point.

To illustrate, we consider three musical events, as in (8) (from Schlenker 2017, 2019a) which are each defined by a pair of two properties, involving chords (and hence harmonic stability), and loudness (in decibels). 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.

(8) 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 an object that gains energy, or one that approaches the perspectival point.

Schlenker 2017, 2019a,b consider different sequences of three events involving two objects, as in (9): a sunrise and a sunset (with the sun gaining or losing luminosity), a boat approaching or departing (with the boat moving closer or further away).

(9) 4 sequences of 3 events (involving the sun or a boat) in a world w relative to a perspectival point π

a. Sun-rise =	sun, <minimal-luminosity, rising-luminosity, maximal-luminosity>
b. Sun-set =	sun, <maximal-luminosity, diminishing-luminosity, minimal-luminosity>
c. Boat-approaching =	boat, <maximal-distance, approach, minimal-distance>
d. Boat-departing =	boat, <minimal-distance, departure, maximal-distance>

In (9)a, the sun rises in 3 stages, starting from one with minimal luminosity from the perspective of π, continuing with one with rising luminosity, and ending with one with maximal luminosity. In (9)b, the sun sets with the opposite movement. In (9)c, a boat approaches in three steps: it is initially at a near-standstill in the distance (its maximal distance in this movement), then it approaches, and then it comes to a stop closer to the perspectival point (we assume that its other properties remain constant). In (9)d, a boat departs with the opposite movement.

In (9)a, the apparent energy of the object rises, as mandated by the Loudness condition; and the first and third events are more stable than the second, 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 greater change). As a result, M as defined in (8) is true of (9)a (relative to the world w and perspectival point π mentioned in (9)). By contrast, a sunset would fail the Loudness condition, as the apparent level of energy of the object does not rise, hence M is false of (9)b. Similarly, interpreting the Loudness condition in terms of proximity rather than in terms of levels of energy, M could be satisfied by a boat approaching, as in (9)c: the boat moves closer to the perspectival point, which correctly interprets the Loudness condition, and the boat is at rest and thus stable at the beginning and end of the movement. By contrast, a boat *departing* could not satisfy the Loudness condition, and thus M is false of (9)d. (A more formal version of this example is discussed in the Appendix in (103); it includes an additional case, involving a car crash.)

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 as a sunset or a boat approaching), and may be very diverse; this is the sense, already noted, in which musical meaning is in general *very* abstract.

On the basis of this bare-bones music semantics, Schlenker 2017, 2019a discusses theoretical extensions pertaining to the interface between (i) syntax and music semantics, and (ii) semantics and pragmatics. They will not be crucial to follow the rest of this piece.

2.3 Testing semantic effects

How should semantic effects be tested in music? There is a long tradition of experiments testing correlations between music and some inferential effects; several are reviewed in Schlenker 2017 in the context of music semantics.[15] Articulated theories of music semantics call for a systematic analysis of

[15] To mention some examples: (i) Eitan and Granot 2006 provide experimental data on the connection between 'inter-onset interval' (= interval between notes) and the scenes evoked in listeners; (ii) the connection between music and movement is discussed in Clarke 2001, Eitan and Granot 2006, Godoy and Leman 2010, Larson, 2012;

minimal pairs, as in linguistics, and while they have so far been based on introspective judgments, they can and should become experimental in the future. As discussed in Schlenker 2017, 2019a, tests start from precise hypotheses, e.g. that all other things being equal, a denoted object will be inferred to be more excited if the music involves a higher-pitched than a lower-pitched sound. The inferences and their triggers should then be assessed by way of minimal pairs, i.e. excerpts that differ just with respect to the crucial property, such as pitch height (several examples of diverse minimal pairs are discussed in the rest of the present piece). Target inferences may be assessed by way of more or less abstract statements in natural language (*Which of these two pieces evokes a phenomenon with the greater level of energy*? *Which of these two pieces best evokes an object moving away?)* or in indirect ways, for instance by having subjects match musical stimuli with non-musical scenes (e.g. visual ones). If one thinks that the inferential mechanism is lifted from auditory cognition, one may in a third step create minimal pairs of non-musical stimuli (e.g. with noise, with human voices, or with animal calls) to test the parameter under study.

A case study was developed by Migotti and Zaradzki 2019. They asked under what conditions a musical snippet could optimally evoke someone walking, and suggested (in part on the basis of how walking works in physical space) that "it has to involve the [1] steady [2] repetition of [3] two different chords", which are "[4] both intrinsically stable, and [5] sufficiently close to each other in the tonal space".[16] They then proceeded to construct minimal pairs that satisfied or violated these five conditions, and asked informants (and then experimental subjects) how well they evoked a person walking. In experimental work, they suggested that the highest endorsements were obtained when all five conditions were met (see also Zaradzki 2021).

In view of the abstract (i.e. underspecified) nature of music semantics, future tests will probably have to discuss less concrete examples, e.g. involving the level of energy of an object, its distance, its animate or inanimate character, etc.

2.4 Objections

Several objections could be raised at this point.[17] First, it is one thing to claim that music *can* be understood to represent something extra-musical, and quite another to claim that it invariably does. To take a point of comparison, the three columns in (1) can but need not be interpreted as a diagram: they may also be perceived as mere geometric shapes, without a semantics. For music, the question is open, and it is complicated by the fact that the semantics is very abstract, which presumably makes some semantic effects less accessible to conscious thought. But we believe that the animated character of music makes it particularly prone to semantic associations (of the type evoked in (3) and in fn. 15, and established in part in the experimental literature).

In a discussion of the expressive power of music, Lerdahl 2001 drew a helpful comparison with Heider and Simmel's (1944) abstract visual 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": they were then perceived as animate agents (video examples [**AV00** https://youtu.be/i3SBv9Xz8zc]). It is hard to resist attributing such intentions to these geometric figures. As Lerdahl argued, 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".[18] We submit that in this case as well one has a strong tendency to perceive

(iii) the implications of loudness, for instance in terms of distance, are discussed in Eitan and Granot 2006 and Ilie and Thompson 2006; (iv) the interpretation of frequency in terms of object size is discussed in Cross and Woodruff 2008; (v) diverse emotional implications of music are discussed in reviews by Gabrielsson and Lindström 2010 and Juslin and Laukka 2003. These implications include for instance association between higher pitch and greater 'tension arousal' (Ilie and Thompson 2006), and between distortion noises and increased arousal and negative valence (Blumstein et al. 2012). For their part, Sievers et al. 2013 find similarities between the mechanisms that trigger emotions in music and in the movement of a ball that can take various shapes, a program further developed in Sievers et al. 2019. See also Koelsch 2012 for a relevant review.

[16] The numbering is ours.

[17] Thanks to two anonymous reviewers for raising the first two objections that follow.

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

sounds as representing something extra-musical (albeit something very abstract). While this is just a conjecture that ought to be tested with experimental means, we will discuss in Section 3.3 several cases in which some abstract semantic implications of musical excerpts are arguably derived rather automatically.[19]

A second foundational question pertains to the fact that, to the extent that music tells stories, these are fictional in nature. One might thus ask how the theory developed in this piece dovetails with issues pertaining to truth in fiction. But here it is wise to divide and conquer: just like language and pictures, music can describe worlds that are merely imagined. This is correctly handled by having a semantics that makes provisions for evaluation in a possible world, as is the case in (5). Whether deep conceptual issues arise because musical stories are fictions is a further question that ought to be treated separately (in the same way, Greenberg 2013, 2019a sets up a formal semantics for pictures without discussing in the same breath the issue of fiction, which is a separate problem).

More pointed objections have also been made to the framework discussed up to this point. Migotti 2019 correctly noticed that interpretation by way of preservation rules akin to those in (6) is too permissive. Consider again (8), 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.

Migotti's criticism has more general validity. In their study of walk-denoting music, Migotti and Zaradzki 2019 (and Zaradzki 2021) observe 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 drawn on the level of energy of the denoted object are determined by those that are in fact physically licensed.[20]

We will now disregard these conceptual and technical 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 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".[21] Another is that music only has an internal semantics, in the sense

[19] We further conjecture that in music performance, interpretive choices that go beyond the musical score are often guided by semantic considerations, i.e. by the kind of abstract 'story' the musician wishes to tell.

[20] Migotti and Zaradzki 2019 also raise a more fundamental issue (just a potential one at this point, as their experimental data do not yet prove its validity). They argue in their 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).

[21] 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).

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.[22] 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

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 "the way it makes you feel when you hear it".[23] 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 intended interpretation.

As briefly noted in Schlenker 2019b (Appendix II), Bernstein's point can 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 narrative, such as the Superman story told by Bernstein. But most strikingly, this story is almost entirely isomorphic to the Don Quixote original. We reproduce in (10) 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.

(10) **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: [**AV01** https://youtu.be/dbGV-gUsEPI] (links from Schlenker 2019b)

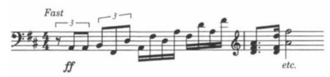
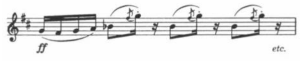
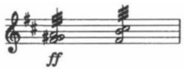
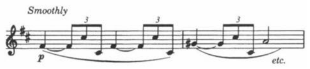
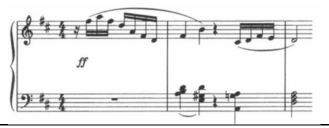
Don Quixote interpretation	**Superman interpretation**	**Salient musical passage**
Context: Don Quixote is a foolish old man who has read too many books about knighthood and decides he is a marvelous knight himself. Sancho Panza is his devoted servant. [**AV01 5:17**[24] https://youtu.be/dbGV-gUsEPI&t=5m17s]	**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. [**AV01 0:28** https://youtu.be/dbGV-gUsEPI&t=28s]	

See also Rodriguez 2021 for a detailed (and historically informed) discussion of music semantics from a philosophical perspectival.

[22] 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).

[23] Special thanks to P. Egré (p.c.) for calling our attention to the relevance of Bernstein's discussion for music semantics. Bernstein revisited this topic in his Harvard Lectures (Bernstein 1976), with different views: "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 (earlier) Young People's Concerts for their rich empirical content and negative thesis rather than for the positive theory Bernstein develops in them.

[24] **AV01 5:17** makes reference to downloadable audiovisual example **AV01** at time 5 minutes 17 seconds; the same notation applies elsewhere in this article.

Don Quixote departs on his horse to conquer the world. [**AV01 5:36** https://youtu.be/dbGV-gUsEPI&t=5m36s]	Superman comes charging along through the alley on his motorcycle. [**AV01 1:08** https://youtu.be/dbGV-gUsEPI&t=1m8s]	
We hear Sancho chuckling to himself.[25] [**AV01 5:45** https://youtu.be/dbGV-gUsEPI&t=5m45s]	Superman whistles his secret whistle (in the woodwinds) so the prisoner will know he's coming. [**AV01 1:20** https://youtu.be/dbGV-gUsEPI&t=1m20s]	
They see a flock of sheep in the field going *baa-baa*. [**AV01 6:03** https://youtu.be/dbGV-gUsEPI&t=6m3s]	Superman hears all the prisoners snoring away peacefully in the dead silence of night. [**AV01 1:28** https://youtu.be/dbGV-gUsEPI&t=1m28s]	
A shepherd is playing on his pipe. [**AV01 1:28** https://youtu.be/dbGV-gUsEPI&t=1m28s]	Superman hears his imprisoned friend playing his kazoo over the snoring, which gets louder as he gets nearer. [**AV01 6:16** https://youtu.be/dbGV-gUsEPI&t=6m16s]	
Don Quixote charges at the sheep, taking them to be an army. [**AV01 6:27** https://youtu.be/dbGV-gUsEPI&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. [**AV01 2:14** https://youtu.be/dbGV-gUsEPI&t=2m14s]	loud bang in the percussion:
The sheep run off in all directions baaing wildly. [**AV01 6:40** https://youtu.be/dbGV-gUsEPI&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. [**AV01 2:22** https://youtu.be/dbGV-gUsEPI&t=2m22s] The snoring gets farther and farther away, until we don't hear it any more. [**AV01 2:37** https://youtu.be/dbGV-gUsEPI&t=2m37s]	
Don Quixote is convinced he has done a truly knightly deed, and is he proud! [**AV01 6:45** https://youtu.be/dbGV-gUsEPI&t=6m45s]	Our hero at last reaches freedom! [**AV01 2:50** https://youtu.be/dbGV-gUsEPI&t=2m50s]	

As illustrated in (11), 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 ((11)a, d, g). The sheep going *baa-baa* corresponds to the prisoners snoring ((11)c, f). And the shepherd playing on his pipe gets reinterpreted as the prisoner playing on his kazoo ((11)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 ((11)b).[26] Structurally, this is virtually the only difference between the two stories.[27]

[25] 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).

[26] In Section 5, we will see another case in which there is some freedom in the choice of the number of objects that are taken to be denoted (this will be cashed out in terms of discourse referents).

[27] As noted in Schlenker 2019b about (11)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?)".

(11) **Correspondence in terms of denoted objects between Bernstein's Don Quixote and Superman interpretations**

Don Quixote interpretation	
a. **Don Quixote** departs on his horse.	**Superman** charges along on his motorcycle.
b. **Sancho chuckles.**	**Superman whistles.**
c. **Sheep** go *baa-baa*.	**Prisoners snore** away peacefully.
d. A **shepherd** pays on his pipe.	The **imprisoned friend** plays 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!

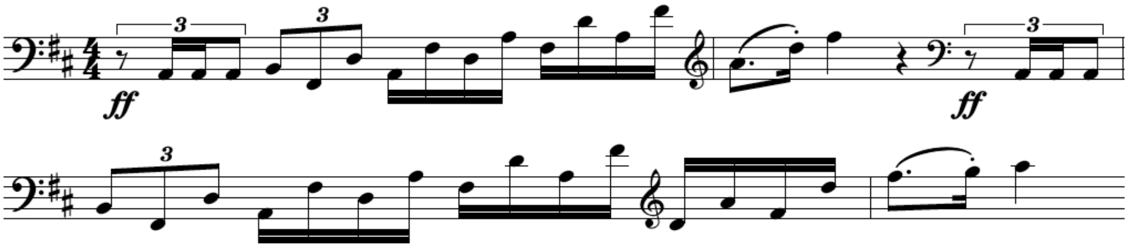
3.3 Music semantics in action: minimal pairs

The point made above 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 the use of a rising frequency to evoke a rise in energy, as in (3)a. This serves to evoke Don Quixote's or Superman's triumphant departure in Bernstein's stories, as shown in (12).

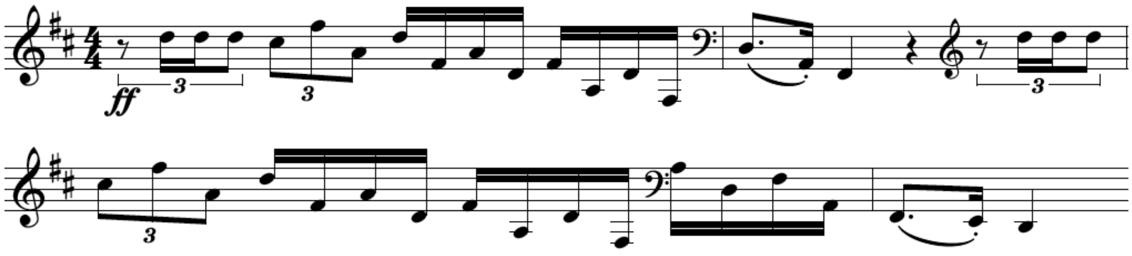
(12) Upwards (original) **[AV18** https://youtu.be/_dSwjTMyzSM**]**



To test the contribution of the rising frequency to the evocation of a triumphant departure, we create (thanks to Arthur Bonetto) a minimal pair that inverses all the melodic motions according to rules of tonal composition.[28] Strikingly, the result is musically acceptable, but the abstract inferences obtained are completely different from those of the original: the general impression of a triumphant departure has been destroyed. This contrast is in part semantic in nature, and our impression is that the triumphant character of the initial piece is very hard *not* to hear, whether one is consciously seeking semantic associations or not. In other words, while abstract, the semantic implications of the music might be derived automatically.

[28] Technically, the recomposition was effected by taking symmetric intervals relative to F# (= the 3rd degree in the relevant key, namely D major). To illustrate, the first note of (12) is A, 2 degrees above F#. The mirror-image note relative to F# going downwards is D, 2 degrees below F#, as in (13). The second note type appearing in (12) is B, 3 degrees above F#. Its mirror-image counterpart relative to F# is C#, 3 degrees below F#, as in (13). This is the reason the second note type appearing in (13) is C#.

(13) Downwards (A. Bonetto) [**AV19** https://youtu.be/5e-39sxEKhk]



3.3.2 Dissonances

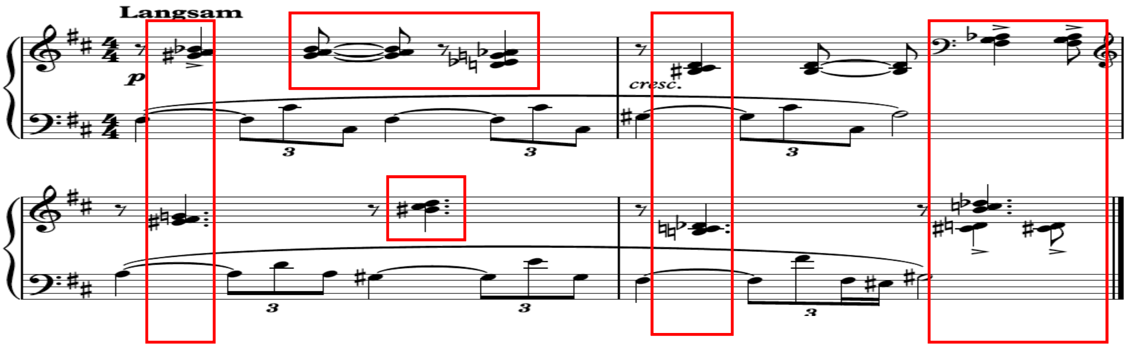
Next, we turn to dissonances used to evoke a flock of sheep going *baa-baa* in the Don Quixote story, and prisoners snoring in the Superman story. As it happens, this is a case in which the musical dissonances are used to evoke chaotic events that are themselves sound-producing, although it can easily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

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

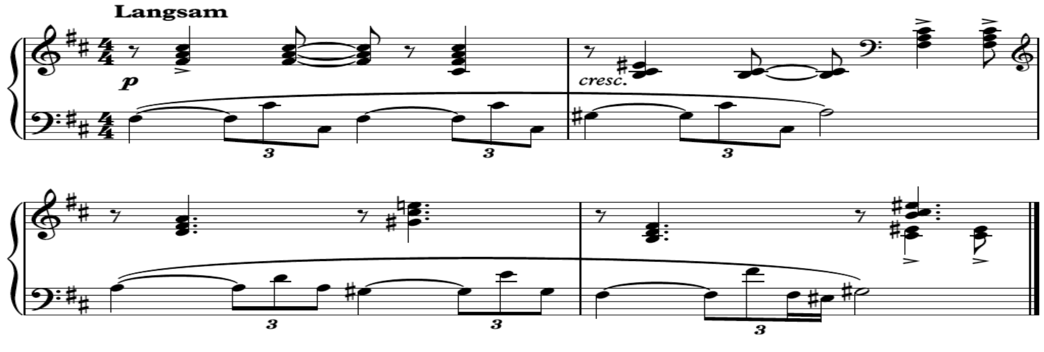
They see a **flock of sheep** in the field going *baa-baa*. [**AV01 6:03** https://youtu.be/dbGV-gUsEPI&t=6m3s]	Superman hears all **the prisoners** snoring away peacefully. [**AV01 1:28** https://youtu.be/dbGV-gUsEPI&t=1m28s]

The dissonances are produced by multiple chords that contain notes that are only one half-tone apart, as shown in the boxed parts of (15). When the music is rewritten so as to minimally remove the dissonances, as in (16), this impression of chaos almost entirely disappears.[29]

(15) Dissonances (original, simplified Midi) [**AV22** https://youtu.be/fKgJDy0wYk0]



(16) No dissonances (A. Bonetto) [**AV23** https://youtu.be/EnhSaeMORCk]



One should not conclude that dissonances in music are solely used to evoke dissonant *sounds* in nature. There are multiple cases in which this is not so. A particularly simple example was mentioned

[29] Bonetto kept the same number of notes in each chord, finding the closest chord that was in the key of the melody (the variation is in D major but this melody is in F# minor).

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 [**AV24** https://youtu.be/6HQqaKEz4tg]). Later in the piece, dissonances are suggestive of the tortoises tripping, as in (17)a. The effect entirely disappears when the music is rewritten so as to remove the dissonances, as in (17)b.

(17) 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 G#'s added (as can be heard by focusing only on the violin and piano parts). [**AV25** https://youtu.be/UqUQQORfCMY]
b. The dissonance can be removed by turning the G#'s into A's – and the impression that tortoises trip disappears (as can be heard by focusing only on the violin and piano parts). [**AV26** https://youtu.be/0A2egp_OlVU]

A dissonance used to evoke an equally silent emotional rather than physical imbalance is used in the music of Hitchcock's *Psycho* [**AV27** https://youtu.be/d3tsfXSJVDs]. The excerpt in (18)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 the *ostinato* repetition of the basic melodic movement and the rhythm), the semantic effect is considerably reduced when the dissonances are removed, as in (18)b,c.

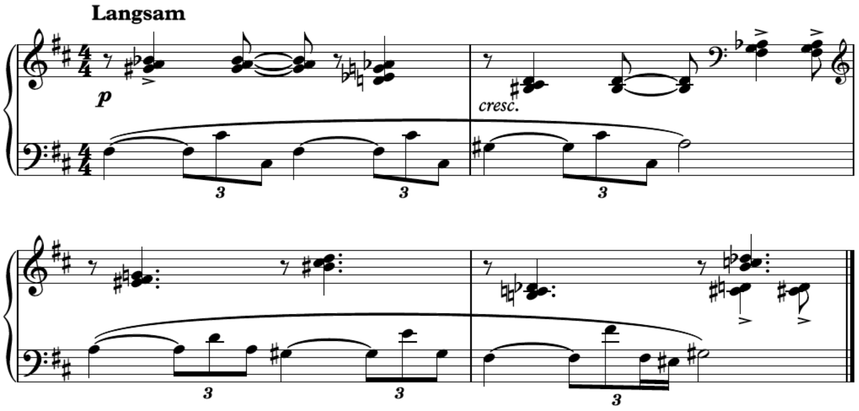
(18) Herrmann's Psycho - reduction, re-written in G minor (A. Bonetto; Schlenker 2019a)[30]
a. Original reduction [**AV28** https://youtu.be/NQUqsiWtI1Q]
b. Same as in a., re-written in G minor without dissonances [**AV29** https://youtu.be/VgNMB9HLXPU]
c. Same as b., closer to the original harmony [**AV30** https://youtu.be/VbdJ_Rbp0a4]

In these cases as well, the impression of instability seems to us to arise whether or not one is explicitly seeking semantic implications: the inferences are arguably triggered automatically.

3.3.3 Loudness

The passage of Strauss's Variation II featuring the sheep going *baa baa* (in (14)) also makes use of a crescendo to indicate that the denoted object is approaching, in accordance with (3)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) Crescendo evoking the sheep (+ shepherd) approaching



Here too, the semantic effect is easy to diagnose by way of minimal pairs. (1)a displays Bernstein's own interpretation in a celebrated performance in 1943.[31] A simplified piano reduction

[30] In greater detail, the transformations were as follows:
(i) From (18)a to (18)b: **Bar 1**: F# > G **Bar 2**: F# > G ; B > Bb **Bars 3-4/6-7** : $\boxed{\text{F > G}}$; Gb > G ; B > Bb **Bar 5:** C > D; B > Bb ; Ab > G ; Eb > D.
(ii) From (18)a to (18)c: same as (i), but the boxed $\boxed{\text{F > G}}$ in (i) becomes $\boxed{\text{F > F\#}}$ instead.
[31] 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

appears in (1)b, also with a crescendo. The same reduction appears in (1)c, but now with a diminuendo (= decreasing loudness) instead of the crescendo. Instead of an impression that something is approaching, we get the impression that something is moving away.[32]

(20) Minimal modifications (A. Bonetto)
a. Dissonances < (Bernstein, 1943) [**AV31** https://youtu.be/sOvqLztu5jo]
b. All <, as in the score (simplified Midi) [**AV32** https://youtu.be/_mMA9dByPAw]
c. All > (simplified Midi) [**AV33** https://youtu.be/kCfay6s4Igs]

This particular excerpt is remarkable in making a clear use of a crescendo to represent an object approaching, but as discussed in earlier work (e.g. Schlenker 2017, 2019a) and in (3)c, there are multiple cases in which loudness modifications provide information about changes in the state of energy or excitement of a denoted object rather than about its distance from the perspectival point.

3.3.4 Cadence

The end of Strauss's Variation II features a cadence evoking a triumphant conclusion. In classical music theory, a (perfect) 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":

> 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 (21); without the ritenuto, the source is easily perceived as moving away.

(21) 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. [**AV36** https://youtu.be/p_52ykHWYJU]
b. In a version of a. without ritenuto, the source seems to be moving away, as it gradually becomes softer, without change of speed. [**AV37** https://youtu.be/bq79oZfHTlk]

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 [**AV38** https://youtu.be/sARBHKoUAkw].

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 role played by the final tonic, as in (23), by replacing it with a cluster, a completely dissonant group of haphazard notes, as in (24).

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."

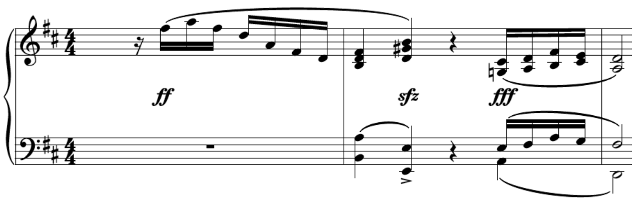
[32] One can also explore more sophisticated minimal pairs in which the melody is played crescendo and the dissonances diminuendo or conversely. The effect is arguably that there are two objects approaching or moving away, as the case may be.

(i) More modifications
[Dissonances <, melody >] (simplified Midi) [**AV34** https://youtu.be/I3vr6p-Tr-4]
[Dissonances >, melody <] (simplified Midi) [**AV35** https://youtu.be/epgwlBhchMY]

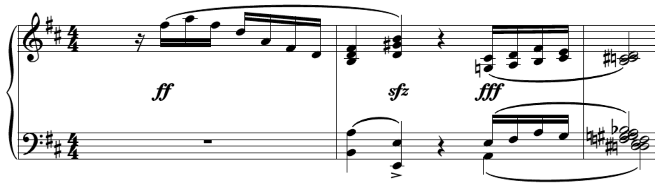
(22) Cadence evoking a triumphant completion

Don Quixote is convinced he has done a truly knightly deed, and is he proud! [**AV01 6:45** https://youtu.be/dbGV-gUsEPI&t=6m45s]	Our hero [= **Superman, with his friend**] at last reaches freedom! [**AV01 2:50** https://youtu.be/dbGV-gUsEPI&t=2m50s]

(23) Expected chord (I) at the end (original) [**AV41** https://youtu.be/eUoyi3Li_ag]



(24) Cluster (A. Bonetto) [**AV42** https://youtu.be/RgppIpK_YYs]



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. And in this case as well, this effect seems to us to be derived automatically, whether or not one is seeking semantic implications in the music.

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 without dissonances; in the same passage, the increasing loudness is naturally interpreted as an object approaching the perspectival point (we can't exclude an alternative interpretation on which it would evoke a rising level of energy of the denoted object, however); and the final, fortissimo cadence is well suited to evoke the attainment of a goal.

Our discussion doesn't do justice to the evocative power of Strauss's music, but it might be enough to show that there is a vast difference between the idea 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 notion 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.[33]

4 Discourse referents: from pictorial semantics to music semantics

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 Abusch 2019, 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

[33] See Schlenker 2019a for a more detailed discussion of the way in which emotions naturally come to play a prominent role in music semantics.

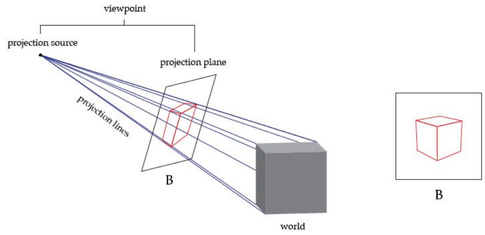
due to Abusch should be borrowed by music semantics. While Greenberg's theory reduced the meaning of pictures to the set of situations that can be projected onto them, Abusch argued that pictures should be enriched with discourse referents akin to those we use in language (for instance to resolve pronominal reference). As we will suggest, the argument can be extended to music.

4.1 Music semantics vs. picture semantics

Following Schlenker 2019b (with small deviations), we can offer a minimal comparison between a Greenbergian semantics for temporally ordered sequences of pictures, and a music semantics. We start from the notion of pictorial truth in (25). The basic intuition is that in a given world and relative to a vewpoing, a picture is true of those eventualities that can project onto the picture, as illustrated in (26) for the case of a system of perspectival projection (since our focus is not on how projections work, we will later leave out the reference to the system of projection).

(25) **Truth-of for a picture (modified from Greenberg 2019a, adding eventualities)**
Let π be a viewpoint, w a world, and S a system of projection. Then:
P is true of eventuality e relative to π and w iff in w e projects to P from π according to S, or in other words: $\text{proj}_S(e, \pi, w) = P$.

(26) **An example of a projection method: perspective projection (Greenberg 2019a)**

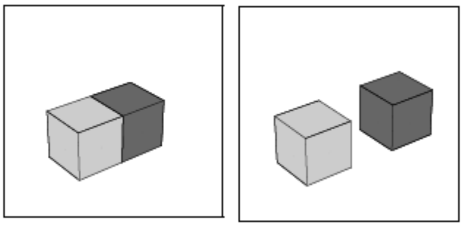


Greenberg 2013, 2019a makes reference to worlds (hence the label *world* next to the cube in (26)), but not to eventualities, which we have added in (25). Eventualities are useful for us to compare pictorial semantics to music semantics. For present purposes, they can be thought of as states of things within a world, akin to situations, hence *parts* of worlds.

What is a viewpoint? For Greenberg (e.g. 2019a), it is "a pair of indices, the first of which gives the spatio-temporal location of the projection source, and the second the spatio-temporal location of the picture plane". Since we will want both components to remain fixed in time (for reasons of simplicity), we will take a Greenbergian viewpoint π to be of the form <π', p>, where π' is a spatial point and p is a projection plane. For terminological simplicity, we will henceforth use the term *perspectival point* to refer to Greenbergian viewpoints, and we will use them in musical and pictorial applications alike. For music, we will only make use of the spatial point coordinate (i.e. π'), disregarding the projection plane (i.e. p). But since we only required (in Section 2.2) that a perspectival point should "include" information about a spatial point, using Greenbergian viewpoints as perspectival points is in keeping with the musical part of our analysis. In effect, we are 'generalizing to the worst case', adding a projection plane for pictorial applications while disregarding it for musical ones. This will be useful when we investigate combinations of music and pictures, as we will need the definition of truth for both media to be relativized to a single perspectival point.

We can then extend this notion of pictorial truth to temporally ordered sequences of pictures, as in the case of the 2-picture sequence in (27), from Abusch and Rooth 2017, which represents "a short comic of two cubes moving apart".

(27) Picture P1 Picture P2



A very simple notion of truth for n temporally ordered pictures can be given as in (28):

(28) **Truth-of for pictorial sequences**
Let π be a viewpoint and w a world. Then:
A pictorial sequence of the form $<P_1, \ldots, P_n>$ is true of eventualities $<e_1, \ldots, e_n>$ relative to π and w iff relative to π and w,
(1) temporally, $e_1 <\ldots< e_n$, and
(2) P_1 is true of e_1 and … and P_n is true of e_n.

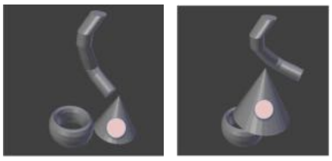
Strikingly, this comes rather close to our definition of truth for musical excerpts in (5), with the difference that the latter definition yielded truth of an object and of a tuple of eventualities, whereas only tuples of eventualities are mentioned in the semantics of pictorial sequences. The difference can be removed by existentially quantifying over objects in the musical case, as in (29) (as we will see below, in pictorial and music semantics alike we must in the end make use of object-denoting variables, and when this adjustment is made, the two semantics will look even more similar; a final comparison can be found in the Appendix).

(29) **Truth-of for musical sequences (modified from (5), with tuples of eventualities only)**
Let π be a perspectival point, and w a world. Then:
A musical sequence $<M_1, \ldots, M_n>$ is true of eventualities $<e_1, \ldots, e_n>$ relative to π and w iff relative to π and w, for some object O, for each k such that $1 \leq k \leq n$, O takes part in e_k and
(1) temporally, $e_1 < \ldots < e_n$;
(2) the Loudness and Harmonic stability conditions are satisfied by O and $<e_1, \ldots, e_n>$ with respect to $<M_1, \ldots, M_n>$.

4.2 Adding discourse referents to pictorial semantics

Abusch 2013, 2019 notices that a definition along the lines of (28) does not do justice to ambiguities that arise in visual narratives, as in the simple example in (30).

(30) **An ambiguity of coreference in pictures** (Abusch 2019)

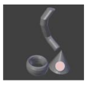
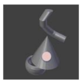


As Abusch 2019 writes, on a simple picture semantics (30)

> 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 (31) suggests that one naturally obtains a reading on which the cross-reference is resolved.[34]

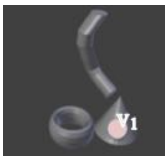
[34] If underspecification is supplemented with a mechanism of enrichment (e.g. pragmatic enrichment), we will end up with an ambiguity of sorts. Our point is just that we plausibly need to generate two readings in this case.

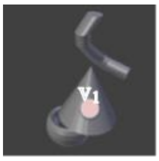
(31) If what happens next is that  , there will be no cone left to close another vase.

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."

How can discourse referents be added to pictorial semantics? In the spirit of Abusch 2013, 2019 (but in the implementation Schlenker 2019b), we can take variables to be distinguished parts of pictures, with a requirement that the object they denote (according to an assignment function) should project onto that part. To be concrete, we can take the part of the picture in which the cone appears to be a variable v_1, with the requirement that the object $g(v_1)$ denoted by v_1 (according to the assignment function g) should in fact project onto that picture part. Having the same variable v_1 appear in (32)a and (32)b will enforce coreference between the two cones.[35]

(32)

a. 

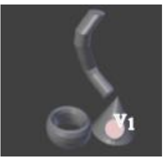
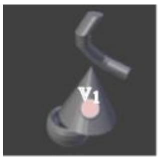
b. 

The definition of truth-of for pictures and pictorial sequences must be refined by taking into account assignment functions, as is done in (33)-(34). The key is the boldfaced requirement in (33), which requires that the objects denoted by the variables take part in the eventualities depicted, and project to the appropriate parts of the pictures.

(33) **Truth-of relative to a perspectival point, a world and an assignment function for individual pictures**
Let π be a perspectival point, w a world, and g an assignment function, and let $P[v_1,\ldots, v_k]$ be a picture containing variables $v_1, \ldots, v_k$. Then:
$P[v_1,\ldots, v_k]$ is true of eventuality e relative to π, w, g iff relative to w and π, e projects to P and **$g(v_1), \ldots, g(v_k)$ are objects that take part in e and respectively project to variables $v_1,\ldots, v_n$ of P**.

(34) **Truth-of relative to a perspectival point, a world and an assignment function for pictorial sequences**
Let π be a perspectival point, w a world, and g an assignment function. Then:
A pictorial sequence of the form $<P_1, \ldots, P_n>$ (where $P_1, \ldots, P_n$ may contain variables) is true of eventualities $<e_1, \ldots, e_n>$ relative to π, w, g iff
(1) temporally, $e_1 <\ldots< e_n$, and
(2) relative to π, w and g, P_1 is true of e_1 and … and P_n is true of e_n.

We can apply these definitions to the two-picture sequence in (32), as in (35). The boldfaced condition enforces coreference between the two cones, as is desired.

(35) For $<P_1, P_2> = <$ ,  $>$, $<P_1, P_2>$ is true of eventualities $<e_1, e_2>$ relative to π, w, g iff
(1) temporally, $e_1 < e_2$, and
(2) relative to π, w and g, P_1 is true of e_1 and P_2 is true of e_2,
iff relative to π and w, (1) temporally, $e_1 < e_2$, and (2) e_1 projects to P_1 and **$g(v_1)$ is an object that takes part in e_1 and projects to variable v_1 of P_1**, and e_2 projects to P_2 and **$g(v_1)$ is an object that takes part in e_2 and projects to variable v_1 of P_2**,

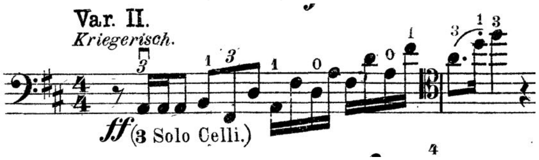
[35] As an anonymous reviewer points out, this means that different picture parts can serve as different tokens of the same variable type. This is the reason we must specify where the variable v_1 appears in each of the pictures in (32).

4.3 Adding discourse referents to music semantics

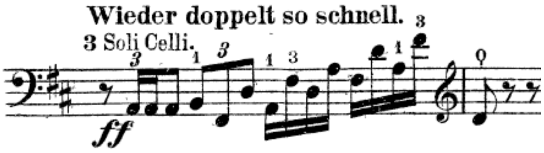
We will now argue that a similar move is warranted in music semantics. To motivate it, let us consider again Strauss's Variation II. The cello melody starts as in (36)a, with the triumphant implications discussed above, corresponding to the beginning of the narration mentioned in (11)a. But within a very different background (pertaining to the dissonances evocative of sheep), almost the same melody appears later, as illustrated in (36)b, when Don Quixote charges at the sheep (a scene described in (11)e). Within Strauss's piece, the intention is clearly that both melodic lines produced by the cello 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 (11)a,e.

(36) Don Quixote departing vs. Don Quixote charging

a. **Variation II, beginning, Don Quixote departing** [**AV43** https://youtu.be/4ETVXSov7xI]



b. **Variation II, Don Quixote charging** [**AV44** https://youtu.be/nUPRnrLEQn8]



Importantly, even in this case the two passages are not strictly identical, as can be seen by the difference in the last bar between (36)a and (36)b: we need a notion of identity that is more abstract that acoustic identity – two excerpts may be acoustically distinct but still count as 'coreferential'. This point is even clearer when we consider larger-scale pieces: throughout Strauss's variations, the cello theme is used to evoke Don Quixote, and acoustic similarity might not be enough to establish these relations of coreference across variations.

In terms of music semantics, it is clear that nothing as specific as a Don Quixote or a Superman story can be evoked, but within each story it is natural to preserve the coreference between (36)a and (36)b. 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 must make decisions about the coreference relations that hold among the discourse referents 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 can be represented with covert discourse referents (i.e. variables). Accordingly, we propose a revised definition of the semantics along the lines in (37), where each musical event is taken to come with a covert variable that corresponds to the object it represents. Cases of ambiguity will arise when one can plausibly decide to coindex two musical events or not to do so.

(37) **Truth-of relative to a perspectival point, a world and an assignment function**
Let π be a perspectival point, w a world, and g an assignment function. Then:
A musical sequence $<M_1[v_1], \dots, M_n[v_n]>$ is true of eventualities $<e_1, \dots, e_n>$ relative to π and g in w iff relative to π, and w, for each k such that $1 \leq k \leq n$, **$g(v_k)$ takes part in e_k** and
(1) temporally, $e_1 < \dots < e_n$;
(2) the Loudness and Harmonic stability conditions are satisfied by $<<g(v_1), e_1>, \dots , <g(v_n), e_n>>$ with respect to $<M_1[v_1] , \dots, M_n[v_n]>$.

The definition in (37) is similar to (5), except that it involves the presence of an object-denoting variable of the form v_k on each musical event. Since music often involves several simultaneous melodic lines or 'voices', it would make sense to have several variables co-occurring at certain times (one for each voice), but this would make the formal system more complicated than is needed for present purposes.[36] As is natural, we require that the objects denoted by the variables should take part in the eventualities that the musical events represent. It will also prove useful in some applications to allow for sums of variables, e.g. v_1+v_2, with the natural interpretation: under an assignment function g, the denotation of the complex variable v_1+v_2 is the mereological sum of the denotation of v_1 and the denotation of v_2, i.e. $g(v_1+v_2) = g(v_1) + g(v_2)$.[37]

The preservation conditions mentioned in (37)(2) should be restated as well: earlier, in (6), we explained under what conditions an object O and some eventualities $<e_1, \ldots, e_n>$ it takes part in may satisfy preservation conditions relative to a musical sequence $<M_1, \ldots, M_n>$. But now different objects may take part in the different eventualities, so we must explain what it means for $<<d_1, e_1>, \ldots , <d_n, e_n>>$ to satisfy preservation conditions with respect to $<M_1, \ldots, M_n>$: instead of considering a sequence of eventualities, we consider a sequences of pairs of an object and an eventuality it takes part in[38], as stated in (38).

(38) **Preservation conditions**
Relative to a perspectival point π and a world w, if for each $i \leq n$ the object d_i takes part in eventuality e_i, a musical sequence $<M_1, \ldots, M_n>$ is true of $<<d_1, e_1>, \ldots , <d_n, e_n>>$ only if $<<d_1, e_1>, \ldots , <d_n, e_n>>$ satisfies the following preservation conditions:
a. Loudness condition
For all $i, k \leq n$, if M_i is less loud than M_k, then in w either:
(i) d_i has less apparent energy from the perspective of π in e_i than d_k does in e_k; or
(ii) d_i is further from π in e_i than d_k is in e_k.

b. Harmonic stability condition
For all $i, k \leq n$, if M_i is less harmonically stable than M_k, then from the perspective of π in w d_i is in a less stable position in e_i than d_k is in e_k.

From (37), we can derive a definition of truth relative to a perspectival point and a world by existentially quantifying over assignment functions and eventualities, as is done in (39).

(39) **Musical truth relative to a perspectival point and a world**
Let π be a perspectival point and w a world. Then:
A musical sequence $<M_1[v_1] , \ldots, M_n[v_n]>$ is true relative to π and w iff for some assignment function g and for some eventualities $e_1, \ldots, e_n$, $<M_1[v_1] , \ldots, M_n[v_n]>$ is true of $<e_1, \ldots, e_n>$ relative to π, w and g.

We will illustrate the role of variables in several derivations later in this piece, but the more urgent matter is to argue for their fruitfulness. In the following discussions, an important intuitive principle will be useful: *all other things being equal,* 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 that in music semantics different timbres give a strong hint as to the non-identity of different denoted objects. We take this to be a default principle that can easily be overridden by other considerations, but it will be useful in the simple examples we will discuss, and it is stated for future reference in (40). (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.)

(40) Default principle: If two musical events involve different timbres, all other things being equal, they tend not to be 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 coindexation.

[36] Allowing for several discourse references per musical time slot would bring music semantics one step closer to the semantics of visual narratives, where several discourse referents may co-occur in one and the same picture.
[37] Technically, we can take an assignment function g to specify a denotation for elementary variables of the form *v*, after which we recursively extend g to complex variables using the rule: *if v+v' is a complex variable, g(v+v') = g(v) + g(v')*. See (96) in the Appendix for a definition along these lines.
[38] The condition that the object takes part in the eventuality is guaranteed by the boldfaced part of (37).

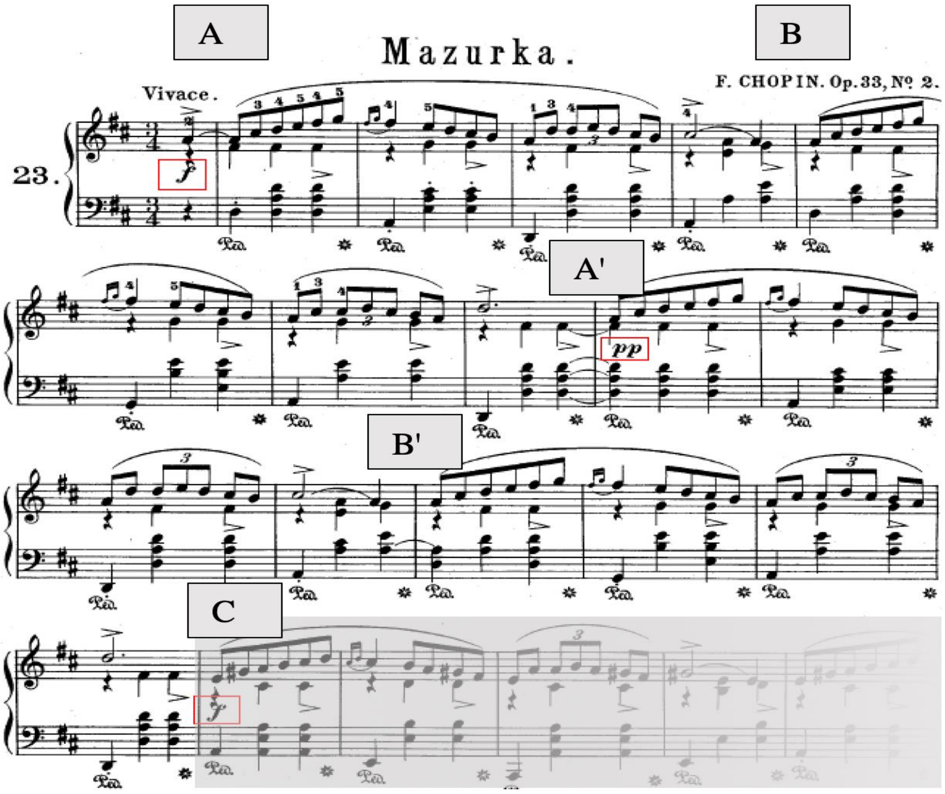
4.4 Revealing discourse referents: an example

To highlight the importance of discourse referents in music semantics, we will investigate a Chopin mazurka[39] in which the music without the dynamics is compatible with different coreference relations among the denoted objects. 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 (40). Finally, the importance of cross-reference relations will become even more salient when we 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.)

4.4.1 Piano version

We consider the beginning of Chopin's Mazurka Op. 33 No 2, reproduced in (41). It has an extremely simple structure of the form AB, repeated, i.e. AB A'B' with A' = A and B' = B (we distinguish the two occurrences of AB because we will need to refer to each occurrence separately). This is part of a broader structure depicted in (42): 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).

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



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

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

We start from a flat realization with constant loudness, as in (43)a. If we assume that there are two denoted objects (which need not be the case), there seem to be at least two salient possibilities: one is that one object corresponds to [AB], while the other corresponds to [A'B']. Alternatively, one object could correspond to A and A', and the other 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 denoted object 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 object is denoted throughout the entire excerpt, but it gets further away, or loses energy, or intentionally repeats itself with less assertiveness in [A'B'].

One can also 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 object, while B and B' correspond to another (as we explain in Section 4.6, 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).

(43) a. Flat realization [AB] [A'B'] [**AV45** https://youtu.be/IMZrU_vEA7A]

b. Chopin's dynamics $[AB]_f$ $[A'B']_{pp}$ [**AV46** https://youtu.be/tyJ0ZQxG8rU]
c. Anti-Chopin dynamics $A_f B_{pp}$ $A'_f B'_{pp}$ [**AV47** https://youtu.be/dYTIAdI4-B0]

As one might expect, most real piano performances follow Chopin's dynamics and thus suggest that, if two denoted objects are present, one corresponds to AB while the other corresponds to A'B' (while probably leaving open the possibility that there is a single denoted object present). An example is Rubinstein's rendition in (44)a. But there are also performances that do not follow Chopin's dynamics. Guller's interpretation in (44)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 object replying to another.

In terms of the formalism introduced in (37), 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 object as AB), but not very compatible with Guller's interpretation (the identity of realization suggests that the same object is involved); this is stated in (45).

(44) Two interpretations of the beginning of Chopin's Mazurka Op. 33 No 2
a. Arthur Rubinstein: $[AB]_f$ $[A'B']_{pp}$ [**AV48** https://youtu.be/Ct7aZ98FfLQ]
b. Youra Guller: $[AB]_{mf}$ $[A'B']_{mf}$ [**AV49** https://youtu.be/R7uY7_Dv6FU]

(45) $[AB]_1$ $[A'B']_2$ with $s(1) \neq s(2)$ is
a. compatible with Rubinstein's interpretation in (44)a
b. not very compatible with Guller's interpretation in (44)b
c. even less compatible with the anti-Chopin dynamics in (43)c (which is suggestive of $[A_1B_2]$ $[A'_1B'_2]$, with $s(1) \neq s(2)$).

Our anti-Chopin dynamics in (43)c tends to suggest that, if two denoted objects are involved, they correspond to A and A' on the one hand and B and B' on the other.

(46) Coreference relations evoked by the anti-Chopin dynamics in (43)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 this difference in denotation to be just a possibility, we will make a note to that effect).

4.4.2 Orchestrations

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 coreference relations among discourse referents (notably because of the principle stated in (40)). We will now consider different orchestrations of the Chopin piece that illustrate this point.[42] We will discuss not just (41) but also the broader environment displayed in (42).

In an orchestration due to Benjamin Britten, the division of labor between instruments is consonant with the division among discourse referents 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 (47) (the orchestra is primarily made of strings).

(47) **Benjamin Britten's orchestration (1941)**[43] [**AV50** https://youtu.be/aHUeA3526DY]
a. [AB]$_{\text{orchestra}}$ [A'B']$_{\text{oboe+flute}}$
b. [CD]$_{\text{orchestra}}$ [C'D']$_{\text{oboe+flute}}$
c. [AB]$_{\text{orchestra}}$ [A'B']$_{\text{oboe+flute}}$

While the Chopin dynamics was consonant with the contraindexing in (45), it did not force it because positing a single denoted object was compatible with the difference in loudness between AB and A'B' (in case that denoted object moved away, lost energy or repeated itself less assertively). By contrast, the association between different timbres and different objects is arguably strong enough that the Britten orchestration suggests contradindexing as in (45), expanded as in (48). In view of the 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 discourse referents are present; they are written as 1 and 2 in (48).[45]

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

[42] 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.

[43] For the history of the Britten score (lost and then rediscovered), see Cooper 2013.

[45] 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** [**AV51** https://youtu.be/2DT-UBh48jA&t=7s]
a. [AB]$_{\text{orchestra}}$ [A'B']$_{\text{winds1}}$
b. [CD]$_{\text{orchestra}}$ [C'D']$_{\text{winds1?}}$
c. [AB]$_{\text{orchestra}}$ [A'B']$_{\text{wind1}}$

(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 denoting a separate object). 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)** [**AV52** https://youtu.be/IMWM8NP3_s4]
a. [AB]$_{\text{orchestra}}$ [A'B']$_{\text{winds1}}$
b. [CD]$_{\text{orchestra}}$ [C'D']$_{\text{winds1+}}$
c. [AB]$_{\text{orchestra+}}$ [A'B']$_{\text{winds1+}}$

(iv) 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 been able to identify), 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 (49). This suggests the coreference relations depicted in (50).

(49) **Other orchestration** [**AV53** https://youtu.be/cabrlIMrq68&t=11m35s]
a. [AB]$_{\text{orchestra}}$ [A'B']$_{\text{winds1}}$
b. [CD]$_{\text{orchestra}}$ [C'D']$_{\text{winds2}}$
c. [AB]$_{\text{orchestra}}$ [A'B']$_{\text{winds1}}$

(50) 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 (51): 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 an object 1 corresponding to the orchestra interacts with three distinct objects 2, 3, 4, as summarized in (52) (we disregard the fact that the orchestra takes slightly different forms in (51)a, b and c).

(51) **Roy Douglas's orchestration** [**AV54** https://youtu.be/1H4ZMhUVafI]
a. [AB]$_{\text{orchestra}}$ [A'B']$_{\text{winds1 oboe bassoon}}$
b. [CD]$_{\text{orchestra}}$ [C'D']$_{\text{winds2 flute}}$
c. [AB]$_{\text{orchestra}}$ [A'B']$_{\text{winds3 clarinet}}$

(52) 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 (53)a,b (where *orchestra+* refers to the enriched orchestra), before the enriched orchestra is used for the entire AB segment in (53)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 variables introduced in Section 4.3, we can represent the coreference relations suggested by this orchestration in (54), where there is overlapping reference among various variables (whether 4 should be taken as a version of the unenriched orchestra isn't entirely clear).[47]

(53) **Similar to Arthur Fiedler**[48] [**AV56** https://youtu.be/srHZ6EtPtD8]
a. [A$_{\text{orchestra}}$ B$_{\text{orchestra+}}$] [A'B']$_{\text{winds1}}$
b. [C$_{\text{orchestra}}$ D$_{\text{orchestra+}}$] [C'D']$_{\text{winds2}}$
c. [AB]$_{\text{orchestra+}}$ [A'B']$_{\text{strings}}$

(54) a. [A$_1$ B$_{1+1'}$] [A'B']$_2$
b. [C$_1$D$_{1+1'}$] [C'D']$_3$
c. [AB]$_{1+1'}$ [A'B']$_{4\text{ (or 1?)}}$

Two conclusions can be drawn. First, through the use of timbre (and the preference rule in (40)), orchestrated music can make coreference indices clearer than piano music. Second, different orchestrations can make difference choices, but not anything goes: Chopin's dynamics (in (45)) suggested that AB and A'B' correspond to different objects, and all the orchestrations discussed here make similar choices in this respect.

[47] 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'.

[48] An orchestration explicitly attributed to Arthur Fiedler is very similar but slightly harder to hear ([**AV55** https://youtu.be/THg_PhplYJk])

4.4.3 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 obtained, there is no requirement that the dance should convey the same information as the music, just in a different modality.[49] Still, by focusing on 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; discourse referents play an important in establishing such bridges between the two media (we will revisit this issue in Section 5).

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. 45). Strikingly, AB in (42) corresponds to a movement by the main ballerina, while A'B' corresponds to a movement by the other dancers, as seen in (55), 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 – see fn. 45): the main ballerina corresponds to object 1, the other dancers (treated as a unique character due to their unity of movement) to object 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 (56).

(55) Fokine's *Les Sylphides* (originally called *Chopiniana)*, movement on Chopin's Mazurka Op. 33 No 2[50]
Performance from 1984, American Ballet Theatre, on an orchestration close to Britten's version
[**AV57** https://youtu.be/Lsuc3KUKKpQ]
$[AB]_{\text{main ballerina}}$ $[A'B']_{\text{other dancers}}$
$[CD]_{\text{main ballerina}}$ $[C'D']_{\text{other dancers}}$
$[AB]_{\text{main ballerina}}$ $[A'B']_{\text{other dancers(+main ballerina)}}$

(56) 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 in music. 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 coindexation. These can be brought out more clearly by orchestration because (all other things being equal) different timbres are naturally associated with different denoted objects. Dance can further highlight the importance of cross-reference by making concrete the identification of certain dancers with certain denoted objects, thus strongly favoring some patterns of cross-reference over others.

4.5 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 (43)c and (46) 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

[49] 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).

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

different timbres are naturally associated with different objects (as stated in (40)). We start from a highly simplified version of Britten's orchestration, in (57)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 (57)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.

(57) a. Simplified version of Britten's orchestration [**AV58** https://youtu.be/UuHU6lXddNc]
$A_{\text{orchestra}}$ $B_{\text{orchestra}}$
$A'_{\text{oboe+flute}}$ $B'_{\text{oboe+flute}}$

b. Anti-Britten (A. Bonetto) [**AV59** https://youtu.be/Y-06MHkX4Uc]
$A_{\text{orchestra}}$ $B_{\text{oboe+flute}}$
$A'_{\text{orchestra}}$ $B'_{\text{oboe+flute}}$

The result is striking: the 'anti-Britten' orchestration in (57)b gives the impression of a dialogue between two entities, whereby A gives rise to the reply in A', and B gives rise to the reply in B'. This is achieved through timbres, which produce (in a stronger way) the coreference relations evoked by the anti-Chopin dynamics in (43)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. 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 denoted objects. A radical example was given in (12) and (13): 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 (57)b. While keeping the coreference relations constant (with one object corresponding to A and A', and another to B and B'), we can modulate the tempo, loudness and articulation to modify the character attributed to each object. In (58)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 (58)b, BB' is played faster, louder and more staccato, and seems more assertive as a result.

(58) a. AA' assertive - BB' fearful (A. Bonetto) [**AV60** https://youtu.be/mIcHD0oYGBc]
AA' (and the piano) is faster, louder, more accented (staccato).
b. AA' fearful - BB' assertive (A. Bonetto) [**AV61** https://youtu.be/zAzF2BLzQtw]
BB' (and the piano) is faster, louder, more accented (staccato).

In sum, the anti-Britten (and anti-Chopin) patterns of coreference explored in (57) (and (43)c) confirm a point already highlighted by the Don Quixote excerpts in (36), and further expanded in the next section: coreference is not reducible to syntactic notions such as grouping. A more abstract semantic notion is needed, corresponding to the identity of the denoted objects. Once this cross-referential device is in place, these objects can be endowed with further individual properties depending on nature and interpretation of the music.

4.6 Syntactic vs. semantic analyses

Coindexation is not the only way to treat two components as a natural unit. Syntactic analyses offer a notion of constituency that might be thought to yield related results (different syntactic theories posit different structures, but the details won't matter here).[51] One might thus think that the unity of AB on the one hand and of A'B' on the other in (57)a should be captured syntactically, by treating AB as a syntactic constituent, and A'B' as another. But on standard analyses, syntactic constituents must be made

[51] In fact, Lerdahl and Jackendoff's pioneering analysis (1983) posited four structures: "Grouping structure describes the listener's segmentation of the music into units such as motives, phrases, and sections. Metrical structure assigns a hierarchy of strong and weak beats. Time-span reduction, the primary link between rhythm and pitch, establishes the relative structural importance of events within the rhythmic units of a piece. Prolongational reduction develops a second hierarchy of events in terms of perceived patterns of tension and relaxation." (Lerdahl 2001)

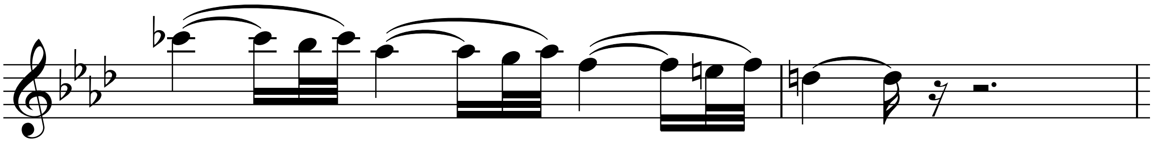
of contiguous elements. This is where the anti-Britten interpretation in (57)b (or for that matter the anti-Chopin dynamics in (43)c) makes an important point, as A and A' need to form a unit, as do B and B'. But A, A' and B, B' are both discontinuous, and neither pair can form a syntactic constituent; by contrast, coindexation has no difficulty dealing with this case. The same situation arose in our initial analysis of the reappearance of the Don Quixote discourse referent in (36) because the excerpts appeared at a considerable distance, making it impossible to treat them as a natural syntactic unit; here too, coindexation seemed to be the correct tool.

Of course one cannot entirely rule out 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.[52] 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.

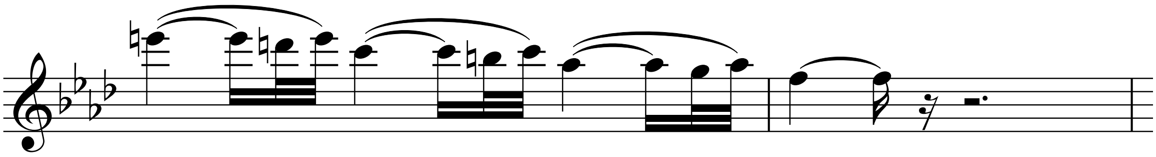
Finally, the very simplicity of the examples we just discussed could give the impression that full acoustic identity is all that's needed to explain coreference in music. But in our initial example in (36), this was not so: the last bar of the cello tune differed across (36)a and (36)b, and yet one had a strong sense that the same object was evoked in the two cases. Similarly, at the beginning of Dukas's Sorcerer's Apprentice, the initial theme is repeated in modified form, first as in (59)a, then higher as in (59)b (higher by one major third), and yet it can easily be understood to involve the very same object.[53] This is in fact the choice that was made in Disney's 1940 version (see (59)c): both occurrences of the theme are associated with the sorcerer. As we will further see in Section 5, although the music may make certain coreference relations particularly natural (as in (59)), there often remain ambiguities of coreference, just as in Abusch's case in (30): positing covert variables seems to be the appropriate tool to represent them.

(59) Initial theme of Dukas's Sorcerer's Apprentice[54]

a. First occurrence: A



b. Second occurrence: A'



c. In Disney's version (1940), both occurrences are associated with the sorcerer [**AV62** https://youtu.be/BR0Asbf2bxg]

4.7 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 coindexation among discourse referents present in the music. These patterns are not reducible to syntactic relations of constituency, which involve contiguity. In music written for a single instrument, such as the piano, coreference relations are typically underspecified, although they are constrained by the dynamics and further properties of the music and interpretation. Orchestration can make coreference relations more explicit, notably when a difference of timbre is indicative of disjoint

[52] See also Lerdahl and Jackendoff 1983 (p. 51) for the role of parallelism in determining how musical elements should be grouped.

[53] In fact, the melodic movement isn't even identical across the two cases: the first movement has a transition Cb-Ab-F, the second E-C-Ab: the last interval of the first movement is a minor third (Ab-F), whereas the last interval of the second is a major third (C-Ab) (thanks to A. Bonetto for discussion).

[54] Thanks to A. Bonetto for transcribing these themes.

reference (as in the orchestrations discussed above). When music is combined with dance and the two share some discourse referents, 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 modifications so as to convey different kinds of information about the denoted objects, 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 Music combined with visual animations

Having brought music semantics and pictorial semantics closer together by adding discourse referents to musical representations, we should ask how the two media interact. One salient case pertains to the dance-music interaction, as we saw in Section 4.4.3, but it is particularly complex: dance presumably has an abstract semantics of its own, but it does not straightforwardly derive from a Greenbergian pictorial semantics (e.g. it is not immediately obvious what a ballet should be a pictorial representation of). We consider instead a simpler case, that of cartoon-music interaction, on the example of the beginning of Disney's Sorcerer's Apprentice (from Disney's Fantasia, 1940).

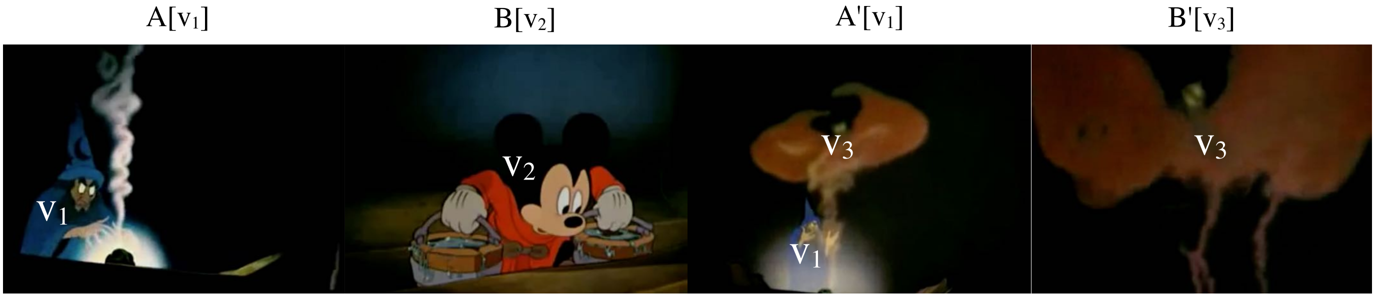
The point we wish to make is simple: the music and the cartoon each have their separate semantics, and thus there is no reason to expect that one should just imitate the other. But the two media jointly can serve to tell a story, with salient points of contact, especially shared discourse referents.

5.1 An example from Disney's Sorcerer's Apprentice

We display in (60) a simplified version of the beginning of Disney's Sorcerer's Apprentice (194). We treat it for simplicity as a sequence of 4 pictures combined with 4 musical events (whereas in fact these are 4 visual scenes combined with 4 musical themes).

The music involves the themes A and A' discussed in (59). A is followed by a very distinct theme B, which re-emerges in modified form (higher) after A', hence a sequence: A B A' B'. The corresponding scenes involve a sorcerer, an apprentice, then a sorcerer again with a genie, and then the genie alone.

(60) **Four pictures from Disney's Sorcerer's Apprentice (below), with four musical events (above) (Disney, Fantasia 1940)**
Video: [**AV62** https ://youtu.be/BR0Asbf2bxg]



In keeping with our analysis, we take the auditory and the musical narrative to each come with variables. The choice among different possible patterns of cross-reference is made to maximize the correspondence between the two media. The correspondence between A and the sorcerer can be enforced by endowing the musical scene and the picture alike with the variable v_1, as is done in (60). The correspondence between B and the apprentice can be enforced by having the variable v_2 in both media. A' features the re-emergence of the sorcerer, hence v_1 again, but the visual scene also has a third character, denoted by v_3, which appears alone in the fourth picture, and can thus be taken to correspond to the object denoted by B'. The music alone would probably suggest that A and A' are coreferential, and that B and B' are as well, and denote a different object from A/A'. But the correspondence with the visual narrative suggests a different pattern of coreference: A and A' are indeed coreferential, but B and B' are not.[55]

5.2 *Truth of mixed sequences*

How should we analyze the semantics of such mixed pictorial and musical sequences? As a first approximation, we take their meaning to be the conjunction of a pictorial and of a musical sequence.[56] Specifically, we take the pictorial and the musical part to characterize the same set of parameters and arguments (perspectival point[57], world, tuple of eventualities), and further connections are imposed by the identity of the variables. The definition of truth-of for mixed sequences is given in (61); it could be turned into a definition of truth *simpliciter* by existentially quantifying the assignment function and the tuple of eventualities.

(61) **Truth-of for mixed sequences**
Let π be a perspectival point, w a world, and g an assignment function. Then:

A pictorial sequence of the form $<P_1, \ldots, P_n>$ (where $P_1, \ldots, P_n$ may contain variables) aligned with a musical sequence $<M_1[v_1], \ldots, M_n[v_n]>$ is true of eventualities $<e_1, \ldots, e_n>$ relative to π, w, g iff $<P_1, \ldots, P_n>$ is true relative to v, w, g and $<M_1[v_1], \ldots, M_n[v_n]>$ is true of $<e_1, \ldots, e_n>$ relative to π, w, g.

5.3 *Application to the example*

A schematic derivation of the truth conditions for the mixed sequence in (60) is provided in the Appendix, where we have assumed for simplicity that the perspectival point remains constant. The details are a bit tedious, but the main results appear in (62) (see (113) for the full derivation of the truth-of conditions, and (114) for truth conditions *simpliciter*).

(62) Let π be a perspectival point, w a world, and g an assignment function. Then:
(60) is true of eventualities $<e_1, e_2, e_3, e_4>$ relative to π, w, g iff
temporally, $e_1 < e_2 < e_3 < e_4$, and relative to π, w, g,
(pictorial component)
e_1 projects to P_1 from π and $\mathbf{g(v_1)}$ takes part in e_1 and projects to variable v_1 of P_1,
and e_2 projects to P_2 from π and $\mathbf{g(v_2)}$ takes part in e_2 and projects to variable v_2 of P_2,
and e_3 projects to P_3 from π and $\mathbf{g(v_1)}$ and $\mathbf{g(v_3)}$ take part in e_3 and respectively project to variables v_1 and v_3 of P_3,
and e_4 projects to P_4 from π and $\mathbf{g(v_3)}$ takes in e_4 and projects to variable v_3 of P_4, *and*
(musical component)
g(v₁) *takes part in e_1 and* ***g(v₂)*** *takes part in e_2 and* ***g(v₁)*** *takes part in e_3 and* ***g(v₃)*** *takes part in e_4, and the Loudness and Harmonic stability conditions are satisfied by $<<\mathbf{g(v_1)}, e_1>, <\mathbf{g(v_2)}, e_2>, <\mathbf{g(v_1)}, e_3>, <\mathbf{g(v_3)}, e_4>>$ with respect to $<A[v_1], B[v_2], A'[v_1], B'[v_3]>$.*

The mechanics works as follows. First, the pictorial sequence and the musical sequence are evaluated with respect to the same perspectival point, world, assignment function and tuple of eventualities: the two media impose their own conditions and are in effect conjoined. Second, the variables v_1, v_2 and v_3 impose patterns of coreference within and across media (the relevant denotations are boldfaced in (62)). Specifically, v_1 enforces coreference between theme A and theme A', which is musically natural in view of their similarity. One would have reason to expect the same pattern of coreference between B and B', but the choice made to endow B with variable v_2 and B' with variable v_3 fails to enforce this. This would be an arbitrary choice if we considered the music alone, but for the mixed sequence this distribution of variables makes good sense. On the pictorial side, just as in Abusch's

[55] Further patterns of indexation could be considered. A. Bonetto (p.c.) suggests that one could take B and B' to be coindexed, but to correspond to something more abstract than the apprentice or the genie. On this interpretation, then, we would have the representations $B[v_4]$, $B'[v_4]$, where v_4 does not corefer with v_1, v_2, v_3.

[56] We write that this is just 'a first approximation' because in some cases one medium is presented as primary and the second as parasitic, in which case the second arguably triggers cosuppositions, as we suggest in Section 6.

[57] Here it should be recalled that we 'generalized to the worst case' by taking perspectival points to be in essence Greenbergian viewpoints, which come with a spatial location (useful for musical and pictorial applications) as well as a projection plane (relevant in the pictorial domain only).

example in (30), variables enforce appropriate patterns of coreference between the sorcerer in the first and in the third pictures, and also between the genie in the third and in the fourth pictures. Finally, the variables establish patterns of cross-reference *across* media, ensuring that themes A and A' are understood to be about the sorcerer, theme B about the apprentice, and theme B' about the genie.

5.4 Derived notions

Having provided a general definition of truth conditions, we can easily obtain derived notions. One is that of *entailment*: a sequence entails another if every set of appropriate parameters and arguments (perspectival points, worlds, assignment functions, tuples of eventualities) that satisfies the first satisfies the second. It is thus immediate that the mixed sequence in (60) entails the pictorial sequence without the music. This is simply because the truth conditions of the latter are given by (62) without the italicized component, and the removal of this conjunct obviously yields a weaker meaning.

Similarly, one can define a notion of *contradiction*: a mixed sequence is contradictory in case no set of appropriate parameters and arguments satisfies it. One particularly interesting case is that in which the music in some way contradicts the pictorial sequence – one may for instance think of a picture of a bomb exploding co-occurring with the most stable cord of a piece (a tonic).

As is standard, entailment and contradiction can be assessed logically, relative to all possible parameters and arguments, or just with respect to those that are compatible with what is assumed in the context. The entailment between (60) and a music-free version of the pictorial sequence holds logically, but some entailments may hold contextually and not logically, and similarly for contradictions.

5.5 Further questions

The present framework could be enriched to apply to more sophisticated cases. First, we have only considered musical excerpts with a single voice, and correspondingly with a single variable. But when several voices are present, several variables should be considered, which would bring music semantics one step closer to the semantics of visual narratives. Second, there could be more complex cases of cross-reference between the music and the visual narrative, especially when several characters appear in the latter. In (60), it made good sense to take the theme A' to be about the sorcerer of the third picture, but were it not for the context (namely the fact that A was about the sorcerer in the first picture), one might have attributed A' to the genie instead. Analyzing these patterns of cross-reference across media could prove insightful to understand mixed sequences (and possibilities will only become richer when several simultaneous voices are considered in an excerpt). Third, we have assumed in our examples that the perspectival point remains constant, but in general this need not be the case.[59] Allowing the perspectival point to shift will also raise interesting questions about the relation between musical and visual perspectives. Fourth, in both music and pictorial semantics, one should in the future explore models with continuous time, which would be far more realistic than the ordered sequences of notes and pictures we discuss in this piece.

The relation between music and dance should be considered within the semantics of mixed sequences, but with important refinements: as noted at the outset, dance presumably has an abstract semantics of its own, one that does not straightforwardly derive from a pictorial semantics, and thus more sophisticated analyses would have to be developed than those entertained in this article. Narrative dances as discussed in pioneering semantic work by Patel-Grosz et al. 2018 might be a particularly good place to start to investigate the dance/music interaction (all the more so since Patel-Grosz et al. specifically argue for the importance of discourse referents in dance).

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

In the previous section, we assumed that in mixed sequences, the music and the pictorial animation are equal partners. This case may indeed arise in music-dance combinations, but with cartoon and film music the visual medium is often viewed as primary, and this arguably has consequences for the status

[59] In Disney's Sorcerer's Apprentice, there is an explicit change of perspectival point between the first and second scene of (60). Our analysis does not capture it. See Abusch and Rooth 2017 and Schlenker 2019a for a discussion of viewpoint shift in narrative sequences.

of the musical enrichment. In brief, we will now argue that, in some cases at least, music can affect the meaning of a cartoon (or of a gif) 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 media. 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 situations a fully well-formed and comprehensible message is preserved when a co-speech gesture is disregarded. Similarly for cartoon and 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, iconic co-speech gestures and 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 music can trigger cosuppositions as well.

6.1 Gestural cosuppositions

While the formal semantic analysis of co-speech gestures and facial expressions is still in its infancy, we will follow recent theories 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). Focusing on iconic gestures that co-occur with propositional and predicative expressions, Schlenker 2018a,b argues that examples such as (63)a(i) differ from at-issue controls like (63)(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*.[60] Questions are an environment in which presuppositions project, as seen in (63)b. The important observation is that the conditionalized inference *if John helps his son, lifting will be involved* projects in the same way in (63)(i) (but not in the control in (63)(ii)).

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

(63) a. Will Ann (i) UP_ **help** her son? (ii) help her son like UP_ **this**?
(i) => If Ann helps her son, she will do so by lifting him

b. Will Ann realize that her son is losing?
=> Ann'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 (64):

(64) a. None of these 10 guys (i) UP **helped** his daughter. (ii) helped his daughter like UP **this**.
(i) => none of these 10 guys helped his daughter; but for each of them, if he had helped his daughter, it would have been by lifting her
b. None of these 10 guys realized that his daughter needed help.
=> for each of these 10 guys, his daughter needed help

Some co-speech facial expressions appear to trigger cosuppositions as well, and certain non-grammatical facial expressions in ASL (American Sign Language) arguably do too (Schlenker 2018a,b). In the English example in (65), conditionalized inferences are obtained in the scope of a question and of a *none*-type quantifier, just as in (63)a(i) and (64)a(i).

[60] 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')".

Notation: :-(is a disgusted facial expression co-occurring with the boldfaced expression.

(65) 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, i.e. yield the same kind of inferences when embedded in diverse logical environments.

One may object that presuppositions ought to be initially accepted by conversation participants, and should thus be uninformative. But in the examples we discussed, co-speech gestures and facial expressions carry non-trivial information, as they drastically modify the meaning of the words they attach to. This objection has not swayed co-speech gesture research because there are numerous cases of informative presuppositions (for discussion, see for instance Stalnaker 2002, von Fintel 2008, Schlenker 2012). To mention but two examples: (66)a can perfectly well be used when the addressee does not initially know that the speaker has a sister, hence the presupposition triggered by the definite description is informative (technically, it forces the addressee to 'accommodate' a context that entails that the speaker has a sister).

(66) a. I can’t come to the meeting – I have to pick up my sister at the airport. (Stalnaker 2002)
b. John's idiotic father has arrived. (Schlenker 2005)
c. ??John's blond father has arrived. (Schlenker 2005)

This is also the case in (66)b, where the possessive description *John's idiotic father* is used in part to inform the addressee that the speaker takes John's father to be an idiot. In fact, the modifier arguably must be informative in this case, for otherwise the denotation could be obtained at lesser cost with the shorter description *John's father*: on Gricean grounds, one expects the modifier to be justified because it is informative, or at least relevant (but see Leffel 2014 for a more fine-grained analysis); this is presumably the reason for the deviance of (66)c.[61]

In sum, although co-speech gestures often make non-trivial contributions, this does not speak against the cosuppositional analysis: many presupposition triggers make non-trivial contributions as well.

6.2 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:

[61] In the context of co-speech gesture theory, Schlenker 2018a (Section 3.3.3) also explores the possibility that gestural cosuppositions put constraints on what the speaker takes for granted (= the speaker's context), rather than on what is common ground in the conversation. This would offer a different reason why cosuppositions need not be trivial at all.

(67)

GLOUP!
GLOUP!
GLOUP!
ZENITRAM

What will happen next: will Asterix… **do what's needed?**[62]

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 (68)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 (68)b.[63].

(68) a. The soldier will not BOOM **[assassinate his target]**. [**AV63** https://youtu.be/W4q5r3hKjvA]
=> if the soldier were to assassinate his target, an explosion would be involved
b. The soldier will not assassinate his target like BOOM this. [**AV64** https://youtu.be/W9bY5j_UXu4]
≠> if the soldier were to assassinate his target, an explosion would be involved
(Pasternak 2019)

In an experimental extension, Pasternak and Tieu 2020 further highlight the similarity between inferences triggered by co-speech gestures and by co-speech sound effects, with very similar inferential patterns across different embeddings under logical operators (*might, not, each, none, exactly one)*

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: in (69), a descending scale co-occurring with the Verb Phrase suggests that if the student were to adjust the brightness, this would involve turning it down.

(69) **The student will not DOWN [adjust the brightness setting of his computer screen].**
[**AV65** https://youtu.be/n-cAoiY3Qns]
=> 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 (70), 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.

(70) [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: [**AV66** https://youtu.be/buwYFvVvt8g]
Speech+music: [**AV67** https://youtu.be/tVOBF-gWwcI]
=> 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, 2021) that an enrichment *p*' that shares a time slot with the main message *p* is presented as being unimportant (e.g. because the addressee's attention is naturally focused on the main message rather than on the enrichment), and that for this reason it should be possible to disregard *p*' without affecting the meaning of the *pp*' combination relative to its context c'. The relevant notion of context is that of a 'local context' as used in presupposition

[62] This picture (which is not by Asterix's creator Uderzo, but by Zenutram) can be found at https://www.deviantart.com/zenitram-anth/art/Asterix-chez-les-freaks-472781613 (retrieved December 9, 2019).
[63] Thanks to Robert Pasternak for authorizing us to cite his sound files.

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 (71) (see Esipova 2019 for a slightly more general statement).[64]

(71) 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, 2021)

If these intuitions are on the right track, one expects to find cosuppositions in further cases in which an enrichment is somehow secondary relative to the main message. In particular, starting from the conjunctive semantics for mixed sequences developed in Section 5, we can add the requirement that the musical part should make a trivial contribution, just like the component *p*' in (61). We will now argue that this in fact happens with film and cartoon music: these may involve musical cosuppositions.

6.3 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 of the film involves an ape playing with and then destroying some bones.

(72) Kubrick's bones scene (from *2001: A Space Odyssey*) [**AV68** https://youtu.be/XXggfYsQYyQ]

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, which retrospectively tells us that the beginning of the scene was far less innocuous than it looked.

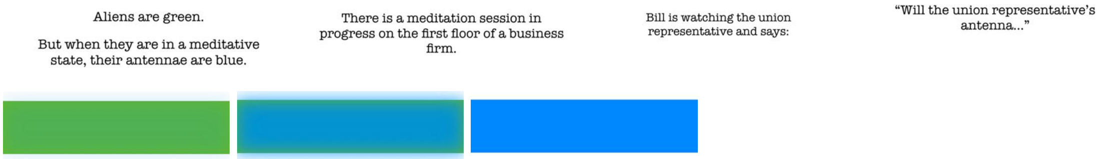
While it is enough to compare the original excerpt with a muted version to see that the music contributes something to the meaning of the scene (more specifically, it infuses it with a momentous or triumphant character), this tells us nothing about the status of this inference: is it at-issue? cosuppositional? something else? Or are these distinctions meaningless for a film? We can begin to address the question by inserting our 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[65] gestures and visual animations') has recently yielded some insights into 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.[66] For instance, in (73), 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 (concretely, from the question in (73) the subjects drew an inference that "the union representative is not currently in a meditative state").

[64] As mentioned at the end of Section 6.1, presuppositions in general and cosuppositions in particular can produce information by forcing the addressee to accept ('accommodate') a context that satisfies the presupposition: the co-speech sound effect in (69) tells the addressee that she is in a context in which adjusting the brightness entails turning it down, and the co-speech music in (70) tells her that for the cavalry to do what's needed, it would have to come quickly and triumphantly.

[65] Just like a co-speech element accompanies a linguistic expression, a pro-speech element *replaces* one.

[66] 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.

(73) **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: [**AV69** https://youtu.be/Nfu9lkLxRxk]



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 combine them with music to test the status of the semantic enrichment contributed by the latter. A first stab is shown in (74). 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 ($^{\text{Zarathustra}}$). 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).

(74) Embedding a Kubrick excerpt in a sentence [**AV70** https://youtu.be/jVhfGjJVZR8]
I saw an ape playing with some bones…
And I wondered…
{Will it… $^{\text{Zarathustra}}$**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 (72) are not at-issue, and are arguably conditionalized in nature. So one would be inclined to endorse the inferences in (75)d and possibly (75)c,e, but certainly not (75)a,b.

(75) **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 for the music to form a coherent whole, as it was inserted by the director so as to have its initial climax a bit later in the film, when the tools are used to kill an animal. In addition, nothing decisively shows 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 semantic import of film music. The examples in (76) were constructed to offer a better coordination between the music and the film excerpt.

(76) **Modifying the Kubrik excerpt, with different musics**
I saw an ape playing with some bones…
And I wondered…
{Will it…
a. $^{\text{Carmen}}$**APE-BREAKING-BONES**} video: Carmen Prelude to Act I (Leonard Slatkin)
[**AV71** https://youtu.be/UnafO0qba4o]
b. $^{\text{Beethoven 8th}}$**APE-BREAKING-BONES**} video: Beethoven 8th, last bars (Kurt Masur)
[**AV72** https://youtu.be/JZ3QWDyKStY]
c. $^{\text{Whistle}}$**APE-BREAKING-BONES**} video: Billy Mowbray Uke and Whistle
[**AV73** https://youtu.be/JHMzDQ6xSzc]

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 (= (76)a), of the triumphant attainment of a goal for the last bars of Beethoven's 8th Symphony (= (76)b), and of something light-hearted for the whistle tune (= (76)c). Rather, our point is that these inferences are not targeted by the question. More specifically, despite the question, one gets conditionalized inferences characteristic of cosuppositions. For the Carmen excerpt, one might thus get the inference that if the ape were to break the bones, this action 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.

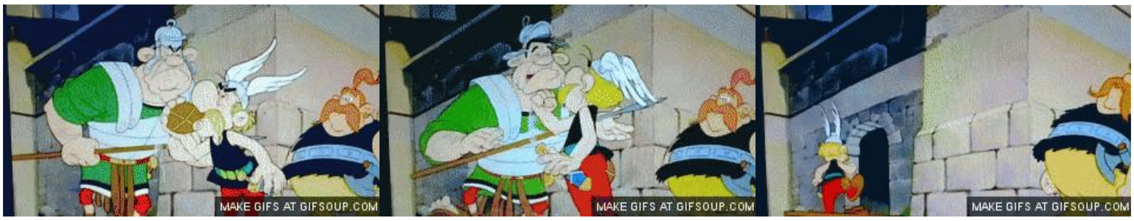
6.4 Refining the argument

Our argument still has one 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 (76). 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 (69)-(70), 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 view. Intuitively, the excerpts discussed in (76) 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 subcomponents of the animation (rather than the global meaning contributed by the gif as a whole).

We start from the gif in (77), which displays Asterix the Gaul drinking a magic potion before hitting a Roman soldier and leaving the premises.[67]

(77) 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 (76)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;[68] thus in (78)a, the tune *WHISTLE* starts as Asterix drinks the potion and continues through the end of the gif, while in (78)a' it starts after Asterix has hit the Roman soldier.

[67] Original from https://giphy.com/gifs/paf-asterix-ulCTAq0E5ekV2, retrieved on December 10, 2019. We modified the original in order 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.

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

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

What will happen next? Will Asterix…

a. ♬ WHISTLE [**AV74** https://youtu.be/aOyr8yTS6uY]

?

a'. ♬ WHISTLE [**AV75** https://youtu.be/BGElapUY4nw]

?

b. ♬ NEWS [**AV76** https://youtu.be/lPkxz9r2cbQ]

?

b'. ♬ NEWS [**AV77** https://youtu.be/I_IPGte_s_g]

?

c. ♬ SUSPENSE [**AV78** https://youtu.be/MvCTK-CYFIo]

?

c'. ♬ SUSPENSE [**AV79** https://youtu.be/MyvFF82lyT4]

?

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 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).[70]

(79) 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 (78)a and (78)a'. In the first case, the light-hearted whistle tune co-occurs with Asterix's entire action, suggesting that the whole sequence would be 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

[70] We also constructed one further pair involving an excerpt from Verdi's Simon Boccanegra , which accompanies a scene in which Simon drinks from a cup which, unbeknownst to him, contains poison (original: [**AV80** https://youtu.be/3mKqInZ1y-Q]): the music suggests that something momentous and disturbing is happening (see Schlenker 2019 for a more detailed semantic discussion). Our consultants did not find the gif-music pairing very successful in this case (the gifs can be seen here: [**AV81a** https://youtu.be/N4-P83l6A-A]; [**AV81b** https://youtu.be/4wV-4_q7ZEM]).

that *if Asterix₂ drinks the magic potion and hits the Roman soldier, then if he leaves the premises thereafter, the latter situation will be light-hearted.*

We believe that related contrasts arise with the other tunes: in (78)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 (78)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 then left, the latter action (leaving) would be triumphant. Similarly, in (78)c an air of mystery is conferred by the music to the entire sequence of events, whereas in (78)c' it is Asterix's departure which is somehow mysterious.

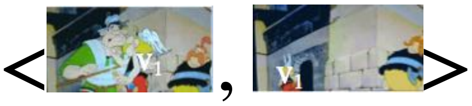
6.5 Analyzing the contrasts

We turn to an analysis of cosuppositions triggered by co-gif music. The technical details are developed in a relatively general form in the Appendix, and here we will just illustrate the main results. To simplify our task, we will analyze a single example, that in which a light-hearted musical event (think of our whistling tune) co-occurs with the second picture, as in (80). We assume that a variable v_1 enforces coreference between the Asterix of the first picture and that of the second (a natural assumption), but also between the Asterix of the pictures and the object the music is about. There are thus three occurrences of the variable v_1: in the first picture, in the second picture, and in the musical tune.

(80) A gif with two pictures <P_1, P_2>, and a light-hearted musical event on the second one

W[v_1]

♬ WHISTLE



P_1 P_2

Inference: if Asterix hits a Roman soldier as shown and then leaves the room as shown, his latter action is light-hearted.

In our official framework, a single musical event couldn't have a non-trivial semantics (because everything is based on preservation conditions, which have some 'bite' for sequences of musical events but not for single ones). But since our goal is to illustrate the workings of cosuppositions, we will make the simplifying assumption that the musical tune (event) is true only of light-hearted eventualities. As explained in the Appendix, we further assume that there can be null musical events (true of everything), which makes it possible to treat the music in (80) as a pair of musical events, the first of which is null and written as Ø, with the assumption in (81):

(81) **Assumption**
<Ø, W[v_1]> is true of <e_1, e_2> from perspectival point π in world w relative to assignment function g iff in w, g(v_1) takes part in e_2 and g(v_1)'s action in e_2 is light-hearted.

In the case at hand, the main thrust of the cosuppositional analysis is that, relative to its local context, the content of the second picture (Asterix is leaving the room) should entail the content of the music (Asterix's action is light-hearted – thanks to the variable v_1 establishing the coreference). But what is the local context of the second picture? It is in essence the global context together with the information provided by the first picture. Technically, it is defined as in (82), which follows from more general considerations discussed in the Appendix: the local context c' of the second picture P_2 is true (relative to a perspectival point π and an assignment function g) of those worlds w and pairs of eventualities <e_1, e_2> for which w is in the global context C, e_1 precedes e_2, and the first picture P_1 is true of e_1.

(82) **Definition of the local context of P_2** (see (115)-(117) for a more general treatment)
Relative to a context set C, the local context c' of P_2 in <P_1, P_2> is defined by:
for each world w, for each perspectival point π, for each assignment function g, for all pairs of eventualities <e_1, e_2>, c' is true of <e_1, e_2> relative to π, w, g iff
(i) w is in C,
(ii) $e_1 < e_2$,
(iii) P_1 is true of e_1 relative to π, w, g.

The next step is to require that, relative to the local context of P_2, the content of P_2 should entail the content of the music. With the local context c' we just defined, this yields the requirement in (83).

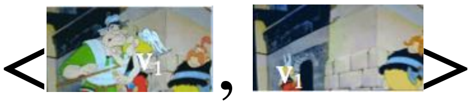
(83) **Requirement that $P_2[v_1]$ entail $W[v_1]$ relative to the local context of P_2** (see (123) for a derivation)
For all worlds w, perspectival points π, assignment functions g, and pairs of eventualities $<e_1, e_2>$ such that: (i) w is in C, (ii) $e_1 < e_2$ and (iii) e_1 projects to P_1 and $g(v_1)$ takes part in e_1 and projects to variable v_1 of P_1,
if relative to π, w, g, e_2 projects to P_2, and $g(v_1)$ takes part in e_2 and projects to variable v_1 of P_2,
then $g(v_1)$'s action in e_2 is light-hearted,
iff
for all worlds w in C, perspectival points π and objects d_1, and pairs of eventualities $<e_1, e_2>$ such that $e_1 < e_2$, e_1 projects to P_1 and d_1 takes part in e_1 and projects to variable v_1 of P_1,
if relative to π, w, e_2 projects to P_2, and d_1 takes part in e_2 and projects to variable v_1 of P_2,
then d_1's action in e_2 is light-hearted.

Informally, we obtain an inference that if Asterix hits a Roman soldier as shown in the first picture and then leaves the room as shown in the second picture, his latter action is light-hearted. This result seems appropriate.

When the music only co-occurs with the first picture P_1, as in (81), the local context of P_1 does not include information about P_2, and as a result the cosuppositional requirement is that if Asterix hits a Roman soldier as shown in the first picture, the latter action is light-hearted (see (121) in the Appendix for a derivation).

(84) A gif with two pictures $<P_1, P_2>$, and a light-hearted musical event on the first one

$W[v_1]$
♬ WHISTLE



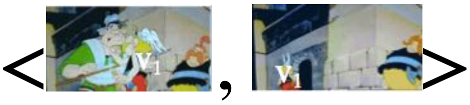
P_1 P_2

Inference: if Asterix hits a Roman soldier as shown, his action is light-hearted.

We also discuss in the Appendix (in (124)) a case that can be entirely treated with the music and pictorial semantics used in this article. It involves two musical events of the same loudness but of different harmonic stability, first a consonant one, $Cons[v_1]$, co-occurring with the first picture, then a dissonant one, $Diss[v_1]$, co-occurring with the second picture, as illustrated in (85).

(85) A gif with two pictures $<P_1, P_2>$, with a consonant chord co-occurring with P_1 and a dissonant chord co-occurring with P_2

$Cons[v_1]$ $Diss[v_1]$



P_1 P_2

When the cosuppositional requirements are computed, we get (among others) an inference that if Asterix hits a Roman soldier as shown and then leaves the room as shown, he is in a less stable position in the latter than in the former eventuality. This, of course, is not at all an inference that would follow from the pictorial sequence alone, which would suggest the opposite.

In sum, on the basis of the conjunctive semantics for mixed sequences developed in Section 5, we can add the requirement that the musical component should make a trivial contribution – presumably because it is in some way parasitic on the pictorial part. Combined with a theory of local contexts for pictorial sequences, this further predicts that musical cosuppositions may enrich the meaning of subcomponents of visual sequences (rather than just the latter as wholes); in view of the examples we discussed, this seems to be a good result.

One point should be added. In our discussion of non-cosuppositional music with pictorial animations in Section 5, we mentioned that a notion of contradiction could be defined. The same result holds when cosuppositions are considered. For instance, if relative to the global context one assumes that the second scene of (85) is more (rather than less) stable than the first, the cosupposition triggered in (85) will lead to a contextual contradiction, with the result that the music could not be made to fit the pictorial animation.

6.6 Limitations and extensions

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 (78), 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).[71] Alternatively, it might be that what matters is the semantic content of the music (e.g. 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[72]) 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.[73]

7 Cosuppositions triggered by music II: pro-speech music

It was recently argued that cosupposition-like inferences are triggered not just by elements that co-occur with the main message, but also by pure iconic elements appearing on their own, 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.

7.1 Cosuppositions triggered by purely iconic elements

Schlenker 2021 argues that ASL classifier predicates (whose movement is entirely iconic although the classifier shape isn't) and some English pro-speech gestures trigger cosuppositions although they do not involve co-speech or co-sign elements. As an example, (86) involves two realizations of a lifting

[71] 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 2021:

(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.

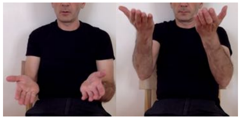
[72] This possibility has been emphasized in work by M. Esipova.

[73] 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: the stimuli are non-linguistic, but they need to be embedded in sentences in order to assess their semantic behavior (e.g. as presuppositions, implicatures, supplements, etc.). The question is whether it is only when they are embedded in linguistic environments that they trigger such inferential types.

gesture: first, a neutral lifting gesture, glossed as *LIFT* in (86)a; second a lifting gesture realized with difficulty (trembling hands), glossed as *LIFT-difficult* in (86)b. The paradigm also includes a gesture-free at-issue control, as in (86)c. Acceptability was rated by three native speakers of American English on a 7-point scale (ratings appear at the beginning of the relevant constructions). The consultants were also asked to assess the strength (on a 7-point scale as well) of the cosuppositional inference: *if the speaker were to lift the child, effort/difficulty would be involved*.

Notation: Acceptability judgments (with 7 = best) appear as superscripts before the examples; inferential strength (with 7 = strongest) follows the examples.

(86) 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

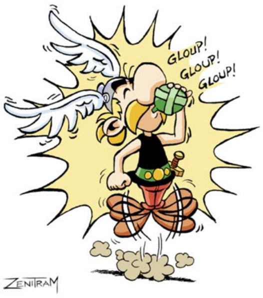
(video 01, a,b,d; 3 consultants; from Schlenker, 2018c and Schlenker 2021)

The results are indicative of a weak cosupposition under questions with *LIFT-difficult* in (86)b but not with the at-issue control in (86)c. Embedding under other operators confirms this pattern of projection (Schlenker 2018c, Schlenker 2021).

We submit that the same effects arise with entirely different iconic forms. If we minimally modify our example from (67) so that the pictorial element now becomes a pro-speech rather than a co-speech picture, as in (87) we obtain a fairly clear meaning: the question is whether Asterix will drink the magic potion.[74]

(87) (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 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 2021. The idea, stated in (88), is that there are two somewhat different reasons why an entailment might be presented as being "unimportant", an undefined notion in (71) above. First, an entailment could be presented as unimportant because it is contributed by a secondary message, one that is parasitic on the main one: this is the case of co-speech and co-sign gestures, and arguably of co-film music. Second,

[74] The picture can be found at https://www.deviantart.com/zenitram-anth/art/Asterix-chez-les-freaks-472781613 (retrieved on December 9, 2019).

however, an entailment could be presented as unimportant for conceptual reasons, for instance because it fails to answer the question under discussion.

(88) 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 2021)

To illustrate, if the question of interest (or "Question under Discussion", e.g. Roberts 1996) is whether Asterix will drink the magic potion, we are faced with the situation illustrated in (89): the question introduces two cells, corresponding to the worlds in which Asterix drinks the magic potion (on the left), and those in which he doesn't (on the right).[75] 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 under Discussion, 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 cell on the left).

[75] Two remarks should be added. First, we assume that the explicit question in (87) serves to address an implicit Question under Discussion: one is really interested in whether Asterix will or will not drink the magic potion. Second, our discussion glosses over the question of how to integrate our pictorial semantics with a compositional semantics for the rest of the sentence (we have not sought to integrate this part of the discussion to the system discussed in the Appendix, as Questions under Discussion introduce numerous complexities of their own).

One way to do things in a simplified case (without *will*) is with the Logical Form in (i). It assumes that, in this linguistic context, the picture has the type-theoretic meaning (ii), which is compatible with the truth-of conditions in (iii). Fixing the perspectival point π, this makes it possible to recover the set of possible worlds that make the sentence true by way of the derivation of truth conditions in (iv) (where we write Asterix' for the denotation of the proper name *Asterix*).

(i) Asterix $\lambda v_1\ \exists e\ P[v_1](e)$

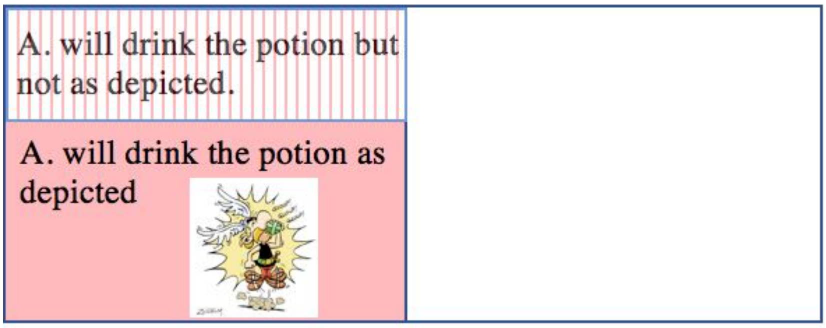
(ii) $[[P[v_1]\ ]]^{\pi, g, w} = \lambda e$ [relative to π, w, g, e projects to P from π and $g(v_1)$ is an object that takes part in e and projects to variable v_1 of P]

(iii) $P[v_1]$ is true of eventuality e relative to π, w, g iff relative to π, w, e projects to P and $g(v_1)$ is an object that takes part in e and projects to variable v_1 of P.

(iv) $[[(i)]]^{\pi, g, w} =\ [[\text{Asterix}\ \lambda v_1\ \exists e\ P[v_1](e)\ ]]^{\pi, g, w} =([[\ \lambda v_1\ \exists e\ P[v_1](e)\ ]]^{\pi, g, w})\ ([[\text{Asterix}]]^{\pi, g, w}) = (\lambda d\ [[\exists e\ P[v_1](e)]]^{\pi, g[v_1\rightarrow d], w})(\text{Asterix}') = [[\exists e\ P[v_1](e)]]^{\pi, g[v_1\rightarrow \text{Asterix}'], w} = 1$ iff for some event e, $[[P[v_1]]]^{\pi, g[v_1\rightarrow \text{Asterix}'], w}(e) = 1$,
iff for some event e, relative to π, w, e projects to P and $g[v_1\rightarrow \text{Asterix}']$ is an object that takes part in e and projects to variable v_1 of P,
iff for some event e, relative to π, w, e projects to P and Asterix' is an object that takes part in e and projects to variable v_1 of P.

(89) **Question under Discussion: Will Asterix drink the magic potion?**

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



A. will drink the potion but not as depicted.

A. will drink the potion as depicted

The generation of a cosupposition is taken to be the minimal pragmatic way to address the problem. Specifically, a presupposition that *if Asterix drinks the magic potion, he will do so as depicted* is exactly what it takes to guarantee that the top-most left-hand cell is ruled out, and thus that a 'no' answer will settle the Question under Discussion; this is shown in (90).

(90) 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 (89), we need a presupposition that we are not in the hatched area, corresponding 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 indeed be the case.

7.2 Cosuppositional effects with pro-speech music?

An initial example of a possible cosuppositional effect triggered by pro-speech music is displayed in (91)b. It features the first phrase of Beethoven's Für Elise (a staple of the student piano repertoire), 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 (91)a, and also with an at-issue control in which the bad interpretation is used with the expression *like this*, as in (92)b.

(91) Which piece will your new student play this afternoon? Will he
a. ELISE ? [**AV82** https://youtu.be/_WF1COwzfhk]
b. ELISE'? [**AV83** https://youtu.be/Cqe7pjbB41U]
b => if he plays, he will do so as shown

(92) Will your new student play like this
a. ELISE ? [**AV82** https://youtu.be/_WF1COwzfhk]
b. ELISE'? [**AV83** https ://youtu.be/Cqe7pjbB41U]
b ≠> if he plays, he will do so as shown

The inference that is arguably derived in (91)b is unsurprising in view of the analysis of pro-speech cosuppositions sketched above: we are presumably interested in which piece the student will play. The literal meaning of the verbal component in (91)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 (89), 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 (93)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, it might already trigger a cosupposition that if the cavalry comes riding in, they will do so skillfully. By contrast, in (93)b the excerpt is played with numerous wrong notes, and this arguably triggers the cosupposition that if the cavalry comes riding in, they will do so in an unskilled fashion. Finally, (93)c features a slow rendition, which becomes even slower as the excerpt progresses, which suggests that if the cavalry comes riding in, they will do so in a slow fashion, and probably with difficulty.

(93) [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 ? [**AV84** https://youtu.be/TOCLGSY-Ntw]
=> if our old cavalry comes riding in, they'll do so skillfully
b. TELL-wrong_notes ? [**AV85** https://youtu.be/1jTNZzGBa48]
=> if our old cavalry comes riding in, they'll do so in an unskilled fashion
c. TELL-slowing_down ? [**AV86** https://youtu.be/i8u0EEUFbuw]
(or: TELL-slowing_down-alternative) [**AV87** https://youtu.be/QONA-MQFaT4]
=> 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.

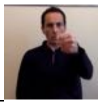
The same disclaimers apply as in our discussion of co-film music. Even if our conclusion is on the right track, it only shows that musical excerpts can trigger cosuppositions *when they are embedded in sentences*. This is compatible with the stronger claim that even without a linguistic environment similar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

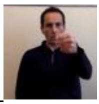
7.3 Can conceptual considerations fully explain cosuppositions triggered by music?

Finally, the existence of cosuppositional effects in pro-speech music forces us to consider one additional possibility for the cosuppositional effects we observed with co-film and co-gif music: could these be due to purely conceptual considerations (combined with questions under discussion), rather than to their parasitic character relative to the main medium?[76]

A similar question arose in co-speech gesture theory: once it was established that some pro-speech gestures can trigger cosuppositions on conceptual grounds (as in (86)b, following (88)(ii)), one could seek to derive *all* cosuppositions in this way. The idea could be that in all cases, some iconic contributions (be they pro-speech or co-speech) provide overly specific information in view of the Question under Discussion, just as was the case with Asterix's manner of drinking in (89): *Asterix will drink the potion as depicted* is just too specific to optimally answer the question *Will Asterix drink the potion?*. A cosupposition (here: *if Asterix drinks the potion, he will do so as depicted*) is exactly what is needed to ensure that both a 'yes' and a 'no' answer address the Question under Discussion. But this unification is implausible: as noted in Schlenker 2021, in (94)a "the co-speech gesture *PUNCH* co-occurs with a VP (*act*, or *do something*, or *take action*) that adds little or nothing to it: it is clear that if one punches, one acts/does something/takes action. Still, a clear cosupposition is triggered, to the effect for instance that *if one took action, this would involve punching the boss*."

(94) [Talking to one's close colleagues]
I am sure tomorrow our boss will once again hurl insults at us, and none of us will

a. PUNCH_ **act / [do something] / [take action]**.

b. PUNCH_.

[76] We thank an anonymous reviewer for pressing us on this issue.

When it comes to co-speech music, this argument is relatively easy to replicate. In Pasternak's example in (69), repeated as (95)a, the descending scale acquires in this context a meaning akin to *turn down,* and it is thus stronger than the modified verb *adjust*, with the result that the latter could be removed will little information loss. This modification is performed in (95)b, where the descending scale is a now pro-speech sound that fully replaces the verb.

(95) a. The student will not DOWN [adjust the brightness setting of his computer screen].
[**AV88** https://youtu.be/idFzcWbdSvI]
=> if the student were to adjust the brightness, this would involve turning it down
b. The brightness setting of his computer screen, the student will not DOWN.
[**AV89** https://youtu.be/vP4Rqr8XeE8]
≠> if the student were to adjust the brightness, this would involve turning it down

Our impression is that (95)a and (95)b both yield an inference akin to: *The brightness setting of his computer screen, the student will not turn down*. But unlike (95)a, (95)b also triggers the cosuppositional inference that *if the student were to adjust the brightness, this would involve turning it down*. Since the informational content between the two verbal expressions is roughly the same (as in both cases, the verb+gesture combination means *turn down*), it is unlikely that conceptual considerations (combined with a Question under Discussion) could suffice to explain the difference. Rather, as in (94)a, the cosupposition seems to be derived because the descending scale is parasitic on the word *adjust*: its co-speech status appears to be crucial to the appearance of the cosupposition.

Developing the same argument for co-film or co-gif music is empirically rather complex, and thus we leave this question for future research.[77]

8 Conclusion

8.1 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". This conclusion does not follow. 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 and Zaradzki 2021 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 should 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 was that the similarity between music semantics and the semantics of pictorial sequences makes it possible to propose an analysis of the meaning of mixed sequences, one in which

[77] A further issue is whether certain contexts could help turn cosuppositions into part of the at-issue component. This is expected on general grounds because presuppositions can be 'locally accommodated' (i.e. turned into the at-issue component) in certain environments, and cosuppositions triggered by co-speech gestures have been argued to be particular prone to this behavior (Schlenker 2018a, Tieu et al. 2017, 2018; see also Esipova 2019).

discourse referents play an important role in establishing relations of coreference within and across media.

Finally, our fourth claim pertained to cases in which music accompanies another medium that can be taken to be primary in the transmission of a message. This is another incarnation of a situation investigated in research on co-speech or co-sign gestures and facial expressions. Recent literature has argued that these gestural expressions trigger cosuppositions, and it has been speculated that this is because they are presented as parasitic on the messages they enrich (both parts are still the subject of intense 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. Going beyond language, we suggested that co-film/gif 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 empirical question for future research is whether cosuppositional effects exist in film or cartoon music that is not embedded in a linguistic environment. On a theoretical level, all analyses offered here should be improved so as to take into account development in time (ordered sequences are just a very rough approximation).[78]

8.2 Broader conclusions

Two further conclusions can be drawn from a broader perspective. First, initial formal attempts emphasized just 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 the existence of discourse referents in visual narratives that led to the exploration of similar issues in music semantics).

Second, we hope that our explorations can 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.

[78] The status of eventualities in our formal analysis should also be clarified. We added them to a mere possible world semantics in order to unify pictorial and music semantics. But we fell short of saying exactly what these eventualities are (besides the claim that they are parts of worlds).

Appendix. Rule and Derivations

We display below the main rules used in this article, and illustrate them with sample derivations. After stating the semantics of complex variables and defining perspectival points (à la Greenberg), we introduce music semantics, pictorial semantics, and their interaction in mixed sequences, first without musical cosuppositions, and then with them.

(96) **Simple and complex variables**
a. Syntax
(i) for any integers i, v_i is an elementary variable;
(ii) if v and v' are variables, v+v' is a complex variable.
b. Semantics
Let g be an assignment of values to elementary variables. In the meta-language, *d+d'* stand for the mereological sum of (possibly plural) objects d and d'. We take g to be defined for elementary variables, and we extend it recursively to complex variables by the following rule:
if *v+v'* is a complex variable, g(v+v') = g(v) + g(v').

(97) **Perspectival points**
A perspectival point π is a pair of the form π = <π', p>, where π' is spatio-temporal point and p is a projection plane. Both π' and p matter in pictorial applications, whereas only π' matters in musical applications.

❑ *Music semantics with variables*

(98) **Musical sequences**
We take musical sequences to be sequences (ordered in time) of musical events, analyzed as tuples of acoustic parameters. In what follows, we simplify things maximally by taking a musical event to be a pair of a harmonic property (a chord written as I, IV, V, etc) and a level of loudness (in decibels), hence for instance: <I, 70db>; <V, 75db>.

(99) **Examples of musical sequences with 3 musical events** (from Schlenker 2017, 2019)
a. M = <<I, 70db>, <V, 75db>, <I, 80db>>
b. M' = <<I, 70db> , <IV, 75db>, <V, 80db>>
c. M" = <<IV, 80db>, <V, 75db>, <I, 70db>>

(100) **Truth-of relative to a perspectival point, a world and an assignment function**
Let π be a perspectival point, w a world, and g an assignment function. Then:
A musical sequence $<M_1[v_1]$, …, $M_n[v_n]>$ is true of eventualities $<e_1$, …, $e_n>$ relative to π and g in w iff relative to π, w, for each k such that $1 \leq k \leq n$, $g(v_k)$ takes part in e_k and
(1) temporally, $e_1 < … < e_n$;
(2) the Loudness and Harmonic stability conditions are satisfied by $<<g(v_1), e_1>$, … , $<g(v_n), e_n>>$ with respect to $<M_1$, …, $M_n>$.

(101) **Preservation conditions**
Relative to a perspectival point π and a world w, if for each $i \leq n$ the object d_i takes part in eventuality e_i, a musical sequence $<M_1$, …, $M_n>$ is true of $<<d_1, e_1>$, … , $<d_n, e_n>>$ only if $<<d_1, e_1>$, … , $<d_n, e_n>>$ satisfies the following preservation conditions with respect to $<M_1$, …, $M_n>$:[79]
a. Loudness condition
For all i, $k \leq n$, if M_i is less loud than M_k, then in w either:
(i) d_i has less apparent energy from the perspective of π in e_i than d_k does in e_k; or
(ii) d_i is further from π in e_i than d_k is in e_k.

b. Harmonic stability condition
For all i, $k \leq n$, if M_i is less harmonically stable than M_k, then from the perspective of π in w d_i is in a less stable position in e_i than d_k is in e_k.

[79] Note that M_1, …, M_n contain variables, but that these do not directly play a role in the preservation conditions. However they do play a role in the underlined part of (100) by requiring that their denotations take part in the appropriate denoted events.

From (100), we can derive a definition of truth in a world relative to a perspectival point (which plays the role of a context) by existentially quantifying over assignment functions and tuples of eventualities, as in (102).

(102) **Truth relative to a perspectival point and a world**
Let π be a perspectival point and w a world. Then:
A musical sequence <$M_1[v_1]$, …, $M_n[v_n]$> is true relative to π in w iff for some assignment function g and for some eventualities e_1, …, e_n, <$M_1[v_1]$, …, $M_n[v_n]$> is true of <e_1, …, e_n> relative to π and g in w.

In (103), we display an example involving a single variable, discussed (without variables) in Schlenker 2017, 2019.

(103) **An example involving a single variable (after Schlenker 2017, 2019)**
a. Musical sequence: M = <<<I, 70db>, v_1>, <<V, 75db>, v_1>, <<I, 80db>, v_1>>
b. Events: we consider the following event sequences, with left-to-right order representing ordering in time, and numerical values assigned to the levels of energy, of proximity and of stability in a world w relative to a perspectival point π

a. Sun-rise	<sun, minimal-luminosity >	<sun, rising-luminosity >	<sun, maximal-luminosity>
Energy	1	2	3
Proximity	1	1	1
Stability	3	1	3

b. Sun-set	<sun, maximal-luminosity >	<sun, diminishing-luminosity>	<sun, minimal-luminosity>
Energy	3	2	1
Proximity	1	1	1
Stability	3	1	3

c. Boat-approaching	<boat, maximal-distance>	<boat, diminishing-distance>	<boat, minimal-distance>
Energy	1	1	1
Proximity	1	2	3
Stability	3	1	3

d. Boat-departing	<boat, minimal-distance>	<boat, rising-distance>	<boat, maximal-distance>
Energy	1	1	1
Proximity	3	2	1
Stability	3	1	3

e. Car-crash	<car, movement_1>	<car, movement_2>	<car, crash>
Energy	1	2	3
Proximity	1	1	1
Stability	3	3	1

(104) **Claim:** Relative to π in w,
a. if $g(v_1)$ = sun, M is true of Sun-rise but not of Sun-set
b. if $g(v_1)$ = boat, M is true of Boat-approach but not of Boat-departing
c. if $g(v_1)$ = car, M is false of Car-crash

Argument:

Ad a.: The Harmonic stability condition is satisfied by both sequences because each has maximally stable events at the beginning and at the end, and a less stable event in the middle, which properly interprets the I-V-I sequence. The Loudness condition is satisfied in Sun-rise since the rising level of energy properly interprets the rising loudness. It is not satisfied in the Sun-set condition: the second chord of M is louder than the first, but this ordering is neither preserved in the 'energy' not in the 'proximity' interpretation in (103)a.

Ad b.: Here too, the Harmonic stability condition is satisfied by both sequences because each has maximally stable events at the beginning and at the end, and a less stable event in the middle, which properly interprets the I-V-I sequence. The Loudness condition is satisfied in Boat-approaching since the increasing proximity properly interprets the rising loudness. It is not satisfied in the Boat-departing condition: the second chord of M is louder than the first, but this ordering is neither preserved in the 'energy' not in the 'proximity' interpretation in (103)b.

Ad c.: The Harmonic stability condition is not satisfied by this sequence because the last chord of M is more stable than the second chord, whereas the last event of Car-crash is less stable than the second event.

Note: we do not derive truth (as opposed to truth-of) conditions because they would be somewhat unilluminating: they are very weak and would just state the existence of a sequence of three eventualities with increasing energy or proximity relative to the perspectival point, and with greater stability at the beginning and at the end than in the middle.

❑ *Pictorial semantics with variables (after Abusch, in the implementation of Schlenker 2019b)*

We turn to a definition of pictorial semantics with variables, taking as primitive the notion of an eventuality projecting onto a picture from a perspectival point in a world (see Greenberg 2013, 2019a for a definition of projection onto a picture from a perspectival point in a world, without reference to eventualities).

(105) **Truth-of relative to a perspectival point, a world and an assignment function for individual pictures**
Let π be a perspectival point, w a world, and g an assignment function, and let $P[v_1,\dots, v_k]$ be a picture containing variables $v_1, \dots, v_k$. Then:
$P[v_1,\dots, v_k]$ is true of eventuality e relative to π, w, g iff relative to π, w, e projects to P and **$g(v_1), \dots, g(v_k)$ are objects that take part in e** and respectively project to variables $v_1,\dots, v_n$ of P.

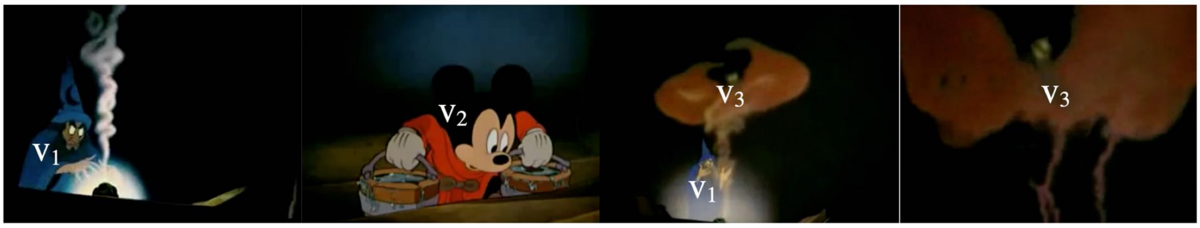
(106) **Truth-of relative to a perspectival point, a world and an assignment function for pictorial sequences**
Let π be a perspectival point, w a world, and g an assignment function. Then:
A pictorial sequence of the form $<P_1, \dots, P_n>$ (where $P_1, \dots, P_n$ may contain variables) is true of eventualities $<e_1, \dots, e_n>$ relative to π, w, g iff relative to π, w and g,
(1) temporally, $e_1 <\dots< e_n$, and
(2) P_1 is true of e_1 and … and P_n is true of e_n.

From (106), we can derive a definition of truth in a world relative to a perspectival point by existentially quantifying over assignment functions and tuples of eventualities, as in (107).

(107) **Truth relative to a perspectival point and a world**
Let π be a perspectival point and w a world. Then:
A pictorial sequence of the form $<P_1, \dots, P_n>$ (where $P_1, \dots, P_n$ may contain variables) is true relative to π, w iff for some assignment function g and for some eventualities $e_1, \dots, e_n$, $<P_1, \dots, P_n>$ is true of $<e_1, \dots, e_n>$ relative to π, w, g.

We turn to an example with four pictures taken from Disney's Fantasia (of course the original has many more pictures, but we simplify maximally for the sake of perspicuity). Here the variables v_1 and v_3 enforce coreference between the sorcerer in pictures P_1 and P_3 and the genie in pictures P_3 and P_4.

(108) **An example: four pictures from Disney's Sorcerer's Apprentice** (Fantasia 1940)



P_1 P_2 P_3 P_4

(109) **Truth-of for (1)**
Let π be a perspectival point, w a world, and g an assignment function. Then for all 4-tuples of eventualities $<e_1, e_2, e_3, e_4>$, $<P_1[v_1], P_2[v_2], P_3[v_1, v_3], P_4[v_3]>$ is true of $<e_1, e_2, e_3, e_4>$ relative to π, w, g iff relative to π, w, g,
(1) temporally, $e_1 < e_2 < e_3 < e_4$, and
(2) $P_1[v_1]$ is true of e_1, $P_2[v_2]$ is true of e_2, $P_3[v_1, v_3]$ is true of e_3, and $P_4[v_3]$ is true of e_4,

iff relative to π, w,
(1) temporally, $e_1 < e_2 < e_3 < e_4$, and

(2) e_1 projects to P_1, and **g(v_1) takes part in e_1 and projects to variable v_1 of P_1**,
and e_2 projects to P_2, and **g(v_2) takes part in e_2 and projects to variable v_2 of P_2**,
and e_3 projects to P_3, and **g(v_1) and g(v_3) take part in e_3 and respectively project to variables v_1 and v_3 of P_3**,
and e_4 projects to P_4, and **g(v_3) takes part in e_4 and projects to variable v_3 of P_4.**

The boldfaced parts enforce the desired coreference between the sorcerer in P_1 and in P_3, and the genie in P_3 and in P_4. In order to obtain truth (rather than truth-of) conditions in a world w relative to a perspectival point π, we existentially quantify the assignment function g and the tuple of events $<e_1, e_2, e_3, e_4>$:

(110) **Truth for (1)**
Let π be a perspectival point and w a world. Then:
$<P_1[v_1], P_2[v_2], P_3[v_1, v_3], P_4[v_3]>$ is true relative to π, w iff for some assignment function g, for some eventualities e_1, e_2, e_3, e_4, relative to π, w,
(1) temporally, $e_1 < e_2 < e_3 < e_4$, and
(2) e_1 projects to P_1, and g(v_1) takes part in e_1 and projects to variable v_1 of P_1,
and e_2 projects to P_2, and g(v_2) takes part in e_2 and projects to variable v_2 of P_2,
and e_3 projects to P_3, and g(v_1) and g(v_3) take part in e_3 and respectively project to variables v_1 and v_3 of P_3,
and e_4 projects to P_4, and g(v_3) takes part in e_4 and projects to variable v_3 of P_4,

iff for some objects d_1, d_2 and d_3, for some events e_1, e_2, e_3, e_4, relative to π, w
(1) temporally, $e_1 < e_2 < e_3 < e_4$, and
(2) e_1 projects to P_1, and **d_1 takes part in e_1** and projects to variable v_1 of P_1,
and e_2 projects to P_2, and **d_2 takes part in e_2** and projects to variable v_2 of P_2,
and e_3 projects to P_3, and **d_1 and d_3 take part in e_3** and respectively project to variables v_1 and v_3 of P_3,
and e_4 projects to P_4, and **d_3 takes in e_4** and projects to variable v_3 of P_4.

The effect of the variables in enforcing coreference relations can once again be seen in the boldfaced parts.

□ *Semantics of pictorial and musical sequences combined*

We turn to the truth conditions of a pictorial sequence of length n aligned with a musical sequence of length n. In essence, such a sequence is true of a sequence of n eventualities just in case the pictorial sequence is true of the n eventualities and the musical sequence is too.

To simplify notations, we will assume that part of a musical or pictorial sequence may be null and thus trivial, in the sense of being true of all eventualities. This makes it possible to only consider the case of n musical events aligned with n pictures. When the music only co-occurs with the end of the pictorial animation, we will take the beginning of the musical sequence to be null, and thus to make a trivial semantic contribution. (The assumption that there could be 'null pictures' will also simplify our discussion of the local context of a picture in a pictorial sequence.)

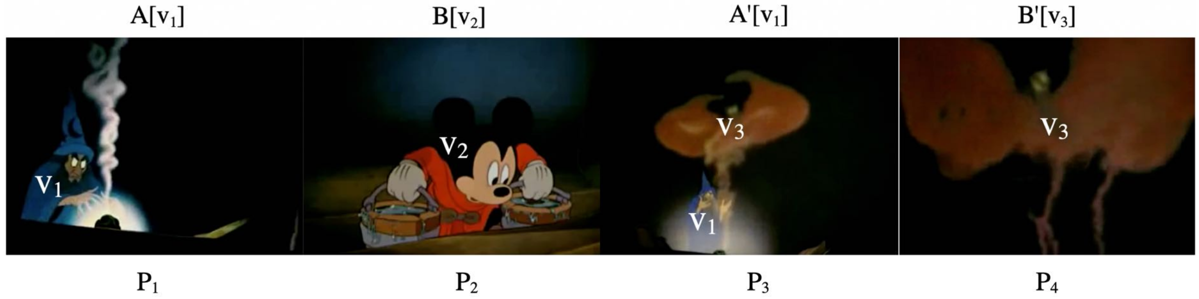
(111) Let π be a perspectival point, w a world, and g an assignment function. Then:
A pictorial sequence of the form $<P_1, \ldots, P_n>$ (where $P_1, \ldots, P_n$ may contain variables) aligned with a musical sequence $<M_1[v_1], \ldots, M_n[v_n]>$ is true of eventualities $<e_1, \ldots, e_n>$ relative to π, w, g iff $<P_1, \ldots, P_n>$ is true relative to π, w, g and $<M_1[v_1], \ldots, M_n[v_n]>$ of $<e_1, \ldots, e_n>$ is true relative to π, w, g.

Notation: We will write as $<P_1, \ldots, P_n> + <M_1[v_1], \ldots, M_n[v_n]>$ a pictorial sequence aligned with a musical sequence.

To illustrate, we consider again the 4-picture sequence in (107), but now aligned with a sequence of four musical events endowed with variables. This is intended as a radical simplification of what happens in the much longer pictorial and musical sequence that appears at the very beginning of Disney's Sorcerer's Apprentice, so we have labeled the four musical events as A, B, A' and B' to evoke the fact that A' is similar to A and B' is similar to B. In keeping with our analysis, the musical sequence is enriched with variables, which disambiguate coreference relations: the notation $<A[v_1], B[v_2], A'[v_1],$

$B'[v_3]>$ indicates that A and A' correspond to the same object, but that B and B', despite their musical similarity, do not. In addition, these variables serve to establish cross-reference between the music and the pictures: as in (107)-(109), v_1 denotes the sorcerer, v_2 denotes the apprentice, and v_3 denotes the genie.

(112) **Four pictures from Disney's Sorcerer's Apprentice (below), with four musical events (above) (Disney, Fantasia 1940)**
[**AV62** https://youtu.be/BR0Asbf2bxg]



(113) **Truth-of for (112)**
Let π be a perspectival point, w a world, and g an assignment function. Then:
$<P_1[v_1], P_2[v_2], P_3[v_1, v_3], P_4[v_3]> + <A[v_1], B[v_2], A'[v_1], B'[v_3]>$ is true of eventualities $<e_1, e_2, e_3, e_4>$ relative to π, w, g
iff $<P_1[v_1], P_2[v_2], P_3[v_1, v_3], P_4[v_3]>$ is true of $<e_1, e_2, e_3, e_4>$ relative to π, w, g and $<A[v_1], B[v_2], A'[v_1], B'[v_3]>$ is true of $<e_1, e_2, e_3, e_4>$ relative to π, w, g,

iff relative to π, w, g,
(pictorial component)
(1) temporally, $e_1 < e_2 < e_3 < e_4$, and
(2) $P_1[v_1]$ is true of e_1, $P_2[v_2]$ is true of e_2, $P_3[v_1, v_3]$ is true of e_3, and $P_4[v_3]$ is true of e_4,
and
(musical component)
$g(v_1)$ takes part in e_1 and $g(v_2)$ takes part in e_2 and $g(v_1)$ takes part in e_3 and $g(v_3)$ takes part in e_4, and
(1) temporally, $e_1 < e_2 < e_3 < e_4$, and
(2) the Loudness and Harmonic stability conditions are satisfied by $<<g(v_1), e_1>, <g(v_2), e_2>, <g(v_1), e_3>, <g(v_3>, e_4>>$ with respect to $<A[v_1], B[v_2], A'[v_1], B'[v_3]>$,

iff relative to π, w, temporally, $e_1 < e_2 < e_3 < e_4$, and
(pictorial component)
e_1 projects to P_1 and $g(v_1)$ takes part in e_1 and projects to variable v_1 of P_1,
and e_2 projects to P_2 and $g(v_2)$ takes part in e_2 and projects to variable v_2 of P_2,
and e_3 projects to P_3 and $g(v_1)$ and $g(v_3)$ take part in e_3 and respectively project to variables v_1 and v_3 of P_3,
and e_4 projects to P_4 and $g(v_3)$ takes in e_4 and projects to variable v_3 of P_4,
(musical component)
$g(v_1)$ takes part in e_1 and $g(v_2)$ takes part in e_2 and $g(v_1)$ takes part in e_3 and $g(v_3)$ takes part in e_4, and the Loudness and Harmonic stability conditions are satisfied by $<<g(v_1), e_1>, <g(v_2), e_2>, <g(v_1), e_3>, <g(v_3>, e_4>>$ with respect to $<A[v_1], B[v_2], A'[v_1], B'[v_3]>$.

Here too, we can obtain truth conditions relative to a world w by existentially quantifying the assignment function g and the tuple of events $<e_1, e_2, e_3, e_4>$:

(114) **Truth for (112)**
Let π be a perspectival point and w a world. Then:
$<P_1[v_1], P_2[v_2], P_3[v_1, v_3], P_4[v_3]> + <A[v_1], B[v_2], A'[v_1], B'[v_3]>$ is true relative to π, w

iff for some assignment function g, for some eventualities e_1, e_2, e_3, e_4, relative to π, w
temporally, $e_1 < e_2 < e_3 < e_4$, and

e_1 projects to P_1, and $g(v_1)$ takes part in e_1 and projects to variable v_1 of P_1,
and e_2 projects to P_2, and $g(v_2)$ takes part in e_2 and projects to variable v_2 of P_2,

and e_3 projects to P_3, and $g(v_1)$ and $g(v_3)$ take part in e_3 and respectively project to variables v_1 and v_3 of P_3,
and e_4 projects to P_4, and $g(v_3)$ takes part in e_4 and projects to variable v_3 of P_4, and

$g(v_1)$ takes part in e_1 and $g(v_2)$ takes part in e_2 and $g(v_1)$ takes part in e_3 and $g(v_3)$ takes part in e_4, and the Loudness and Harmonic stability conditions are satisfied by $<<g(v_1), e_1>, <g(v_2), e_2>, <g(v_1), e_3>, <g(v_3>, e_4>>$ with respect to $<A[v_1], B[v_2], A'[v_1], B'[v_3]>$,

iff for some objects d_1, d_2, d_3, for some eventualities e_1, e_2, e_3, e_4, with respect to π, w, temporally, $e_1 < e_2 < e_3 < e_4$, and

e_1 projects to P_1, and d_1 takes part in e_1 and projects to variable v_1 of P_1,
and e_2 projects to P_2, and d_2 takes part in e_2 and projects to variable v_2 of P_2,
and e_3 projects to P_3 from π and d_1 and d_3 take part in e_3 and respectively project to variables v_1 and v_3 of P_3,
and e_4 projects to P_4 from π and d_3 takes part in e_4 and projects to variable v_3 of P_4, and

d_1 takes part in e_1 and d_2 takes part in e_2 and d_1 takes part in e_3 and d_3 takes part in e_4, and the Loudness and Harmonic stability conditions are satisfied by $<<d_1, e_1>, <d_2, e_2>, <d_1, e_3>, <d_3, e_4>>$ with respect to $<A[v_1], B[v_2], A'[v_1], B'[v_3]>$.

❑ *Local context computation in pictorial sequences*

In order to compute cosuppositions triggered by music co-occurring with a pictorial sequence, we must compute the local context of the beginning of a pictorial sequence (possibly enriched with music). The reason is that presuppositions in general and cosuppositions in particular are conditions that must be satisfied in a local context. As is standard (Heim 1983, Schlenker 2009, 2010), the local context of an expression *E* is computed from the global context C (seen as a set of possible worlds compatible with what is taken for granted in the conversation) together with the expressions that appear before *E*.

We assume that we know the length n of a pictorial sequence $<P_1, \ldots, P_n>$ (where $P_1, \ldots, P_n$ may contain variables), and ask what is the local context of the i^{th} picture P_i. As in Schlenker 2009, 2010, we take this local context to be (the value of) the strongest conjunct one can add to the target expression $<P_1, \ldots, P_{i-1}, \ldots>$ in such a way that, no matter how it ends (starting with a picture P'_i in position i), the truth conditions will not be modified relative to the global context C. We can view a picture sequence of length n as a predicate of perspectival points, worlds, assignment functions and n-tuples of eventualities. This leads to one possible definition of the local context of a picture in a pictorial sequence:[80]

(115) **A possible definition of the local context of a picture in a pictorial sequence of length n**
In a pictorial sequence $<P_1, \ldots, P_{i-1}, P_i, \ldots, P_n>$, the local context of P_i ($1 \leq i \leq n$) relative to a context set C is the strongest c' (true of perspectival points, worlds, assignment functions and n-tuples of eventualities)[81] such that for all $P'_i, \ldots, P'_n$, for each perspectival point π, for each w in C, for each assignment function g, for all n-tuples of eventualities $<e_1, \ldots, e_n>$,
$<P_1, \ldots, P_{i-1}, P'_i, \ldots, P'_n>$ is true of $<e_1, \ldots, e_n>$ relative to π, w, g iff $<P_1, \ldots, P_{i-1}, P'_i, \ldots, P'_n>$ **and c'** are true of $<e_1, \ldots, e_n>$ relative to π, w, g.
Note: For i = 1, the requirement is naturally that for all $P'_1, \ldots, P'_n$,
$<P'_1, \ldots, P'_n>$ is true of $<e_1, \ldots, e_n>$ relative to π, w, g iff $<P'_1, \ldots, P'_n>$ **and c'** are true of $<e_1, \ldots, e_n>$ relative to π, w, g.

(116) **Claim:** Relative to a context set C, the local context c' of P_i in $<P_1, \ldots, P_{i-1}, P_i, \ldots, P_n>$ is defined by:
for each world w, for each perspectival point π, for each assignment function g, for all n-tuples of eventualities $<e_1, \ldots, e_n>$, c' is true of $<e_1, \ldots, e_n>$ relative to π, w, g iff relative to π, w, g
(i) w is in C,

[80] We write 'one possible definition' because others could be explored. For example, we have taken the global context set C to be a set of possible worlds, but it would make sense to take it to be a set of tuples, e.g. including a perspectival point and a possible world, or even a perspectival point, a possible world and an assignment function.
[81] Technically, a local context is a function from such parameters to truth values.

(ii) $e_1 < \ldots < e_{i-1} < e_i < \ldots < e_n$,
(iii) for all k such that $1 \leq k \leq i\text{-}1$, P_k is true of e_k.
Note: For i = 1, condition (iii) is vacuous and the requirement boils down to conditions (i) and (ii).

(117) **Proof:** We need to show that (i) c' as defined in (116) satisfies the equivalence in (115), and that (ii) no stronger c" does (both parts are needed to show that c' is the *strongest* such element, as required). (i) is immediate, as the contribution of c' as stated doesn't add anything to the beginning (up to and including P_{i-1}) of the picture sequence. Part (ii) follows from the assumption that there can be null pictures. Assume that some c" is false of some set of parameters of which c' is true. We will show that c" falsifies the equivalence in (115), showing that c' is in fact the strongest element that satisfies the equivalence.

Suppose, then, that for some world w, for some perspectival point π, for some assignment function g, for some n-tuple of eventualities $<e_1, \ldots, e_n>$,
(i) w is in C,
(ii) $e_1 < \ldots < e_{i-1} < e_i < \ldots < e_n$,
(iii) for all k such that $1 \leq k \leq i\text{-}1$, P_k is true of e_k relative to π, w, g,[82]
but c" is false of $<e_1, \ldots, e_n>$ relative to π, w, g. Take $P_i, \ldots, P_n$ to all be null, which we will write as $Ø_i, \ldots Ø_n$. Now $<P_1, \ldots, P_{i-1}, Ø_i, \ldots Ø_n>$ is true of $<e_1, \ldots, e_{i-1}, e_i, \ldots, e_n>$ relative to π, w, g, but c" is false of $<e_1, \ldots, e_{i-1}, e_i, \ldots, e_n>$ relative to π, w, g. This falsifies the equivalence in (115), as desired.

- ❑ *Musical cosuppositions*

Cosuppositional requirements for musical sequences aligned with pictorial sequences could be computed by requiring that, relative to the global context (i.e. to the context set), the content of the pictorial sequence entail the content of the music. But this would fail to take into account an important asymmetry: if a musical event co-occurs with the beginning of a 2-picture sequence, as in (118), the cosuppositional requirement is arguably that if the first event happens, it should satisfy the content of the music, not that if the entire sequence of events unfolds, the first event will satisfy the content of the music. Concretely: from (118), with a light-hearted whistling co-occurring with P1, we infer that if Asterix punches a Roman soldier, this event will be light-hearted, rather than: if Asterix punches a Roman soldier and then leaves the room, the first event will be light-hearted. The latter requirement makes little sense, since as we watch and hear the mixed sequence, we do not yet know what the end of the pictorial sequence will be, and thus we draw inferences online, on the basis of the beginning of the sequence alone.

(118) A gif with two pictures $<P_1, P_2>$, and a light-hearted musical event on the first one

W[v_1]
♬WHISTLE

< >

P_1 P_2

These considerations show that we need to appeal to local contexts (rather than just to global contexts) as we compute cosuppositional requirements. The initial idea is that relative to a local context, the semantic contribution of a picture P_i to a pictorial sequence should entail the semantic contribution of a musical event M_i to the corresponding (aligned) musical sequence. But in the present framework it makes little sense to talk of the content of a single musical event (because preservation conditions pertain to entire musical sequences, not to individual musical events). So we will state instead that, if a sequence of eventualities $<e_1, \ldots, e_i, \ldots>$ satisfies the local context of P_i, then if P_i is true of e_i, the music should be true of some extension of the sequence $<e_1, \ldots, e_i>$.

(119) **A possible definition of a cosuppositional requirement for mixed sequences**
Consider a pictorial sequence $<P_1, \ldots, P_n>$ (where $P_1, \ldots, P_n$ may contain variables) aligned with a musical sequence $<M_1[v_1], \ldots, M_n[v_n]>$, and let C be a context set.
Then for each i such that $1 \leq i \leq n$, the local context c_i of P_i should guarantee that:
for all worlds w, perspectival points π, assignment functions g, and n-tuples of eventualities $<e_1, \ldots, e_i, e_{i+1}, \ldots, e_n>$, if c_i is true of $<e_1, \ldots, e_n>$ relative to π, w, g, then:

[82] For i = 1, condition (iii) is vacuous.

if P_i is true of e_i relative to π, w, g, then for some $e'_{i+1}, \ldots, e'_n$, $<M_1[v_1], \ldots, M_n[v_n]>$ is true of $<e_1, \ldots, e_i, e'_{i+1}, \ldots, e'_n>$ relative to π, w, g.

We are now in a position to illustrate the cosuppositional requirement on the example of (118). We will simplify the discussion by making the assumptions in (120).

(120) **Assumptions**
a. $<\emptyset, W[v_1]>$ is true of $<e_1, e_2>$ from perspectival point π in world w relative to assignment function g iff in w $g(v_1)$ takes part in e_2 and in w $g(v_1)$'s action in e_2 is light-hearted.
b. $<W[v_1], \emptyset>$ is true of $<e_1, e_2>$ from perspectival point π in world w relative to assignment function g iff in w $g(v_1)$ takes part in e_1 and in w $g(v_1)$'s action in e_1 is light-hearted.

(121) **Cosuppositional requirement in (118)**
a. **Local context**
The local context c_1 of P_1 is defined by:
c_1 is true of $<e_1, e_2>$ relative to π, w, g iff relative to π, w,
(i) w is in C,
(ii) $e_1 < e_2$.
b. **Cosupposition**
For all worlds w, perspectival points π, assignment functions g, and pairs of eventualities $<e_1, e_2>$, if c_1 is true of $<e_1, e_2>$ relative to π, w g, then:
if P_1 is true of e_1 relative to π, w, g, then for some e'_2, $<W[v_1], \emptyset>$ is true of $<e_1, e'_2>$ relative to π, w, g,

iff for all worlds w in C, perspectival points π, assignment functions g, and pairs of eventualities $<e_1, e_2>$, if relative to π, w, $e_1 < e_2$, and e_1 projects to P_1, and $g(v_1)$ takes part in e_1 and projects to variable v_1 of P_1, then for some e'_2, $<W[v_1], \emptyset>$ is true of $<e_1, e'_2>$ relative to π, w, g

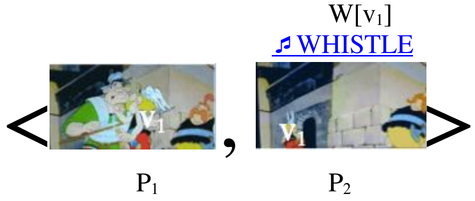
iff for all worlds w in C, perspectival points π, assignment functions g, and pairs of eventualities $<e_1, e_2>$, if relative to π, w, $e_1 < e_2$ and e_1 projects to P_1, and $g(v_1)$ takes part in e_1 and projects to variable v_1 of P_1, then $g(v_1)$'s action in e_1 is light-hearted.

iff for all worlds w in C, perspectival points π, objects d_1 and pairs of eventualities $<e_1, e_2>$, if relative to π, w, $e_1 < e_2$, and e_1 projects to P_1, d_1 takes part in e_1 and projects to variable v_1 of P_1, then d_1's action in e_1 is light-hearted.

Informally: if Asterix hits a Roman soldier as shown, his action is light-hearted.

We turn to the case in which the light-hearted musical event co-occurs with the second picture, as in (122).

(122) A gif with two pictures $<P_1, P_2>$, and a light-hearted musical event on the second one



(123) **Cosuppositional requirement in (122)**
a. **Local context**
The local context c_2 of P_2 is defined by:
c_2 is true of $<e_1, e_2>$ relative to π, w, g iff relative to π, w, g,
(i) w is in C,
(ii) $e_1 < e_2$,
(iii) P_1 is true of e_1, i.e. e_1 projects to P_1, and $g(v_1)$ takes part in e_1 and projects to variable v_1 of P_1.
b. **Cosupposition**
For all worlds w, perspectival points π, assignment functions g, and pairs of eventualities $<e_1, e_2>$, if c_2 is true of $<e_1, e_2>$ relative to π, w, g, then:
if P_2 is true of e_2 relative to π, w, g, then $<\emptyset, W[v_1]>$ is true of $<e_1, e_2>$ relative to π, w, g,

iff for all worlds w in C, perspectival points π, assignment functions g, and pairs of eventualities $<e_1, e_2>$ such that relative to π, w, $e_1 < e_2$, and e_1 projects to P_1, and $g(v_1)$ takes part in e_1 and projects to variable v_1 of P_1,

if relative to π, w, e_2 projects to P_2, and $g(v_1)$ takes part in e_2 and projects to variable v_1 of P_2,
then <Ø, W[v_1]> is true of <e_1, e_2> relative to π, w, g,

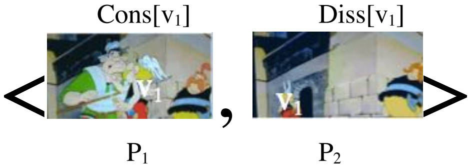
iff for all worlds w in C, perspectival points π, assignment functions g, and pairs of eventualities <e_1, e_2> such that relative to π, w, $e_1 < e_2$, and e_1 projects to P_1, and $g(v_1)$ takes part in e_1 and projects to variable v_1 of P_1,
if relative to π, w, e_2 projects to P_2, and $g(v_1)$ takes part in e_2 and projects to variable v_1 of P_2,
then $g(v_1)$'s action in e_2 is light-hearted,

iff for all worlds w in C, perspectival points π, objects d_1, and pairs of eventualities <e_1, e_2> such that relative to π, w, $e_1 < e_2$, and e_1 projects to P_1, and d_1 takes part in e_1 and projects to variable v_1 of P_1,
if relative to π, w, e_2 projects to P_2, and d_1 takes part in e_2 and projects to variable v_1 of P_2,
then d_1's action in e_2 is light-hearted.

Informally: if Asterix hits a Roman soldier as shown and then leaves the room as shown, his latter action is light-hearted.

Within the musical framework developed in this piece (i.e. without stipulating, as in (120), the semantic content of a musical sequence), we can consider a case in which the same picture pair co-occurs with two chords, first a consonant one written as *Cons*, then a dissonant one written as *Diss*. Keeping the loudness constant, we may for instance think of the succession <I, 70db>; <cluster, 70db> (where a cluster is a chord comprising at least three adjacent notes separated by a half-tone, hence a highly dissonant chord). As stated, the preservation conditions in (38) impose no requirement arising from loudness (because the two chords have the same loudness), but the harmonic stability of chords is interpreted: in essence, the musical sequence will be true of pairs of eventualities <e_1, e_2> that share a participant d who is in a less stable position in e_2 than in e_1 (this is not at all what the pictorial sequence on its own would suggest, so the music will make a non-trivial contribution, possibly even one that contradicts the pictorial content).

(124) A gif with two pictures <P_1, P_2>, with a consonant chord co-occurring with P_1 and a dissonant chord co-occurring with P_2



(125) Let π be a perspectival point, w a world, and g an assignment function. Then:
<Cons[v_1] , Diss[v_1]> is true of eventualities <e_1, e_2> relative to π, w, g iff relative to π, w, $g(v_1)$ takes part in e_1 and e_2, and
(1) temporally, $e_1 < e_2$;
(2) $g(v_1)$ is in a less stable position in e_2 than in e_1.

We can now compute the cosuppositional requirements on P_1 and P_2.

(126) **Cosuppositional requirement on P_1 in (124)**
(starting with a counterpart of the underlined part of the derivation in (121))
For all worlds w in C, perspectival points π, assignment functions g, and pairs of eventualities <e_1, e_2>, such that,
if relative to π, w, $e_1 < e_2$, and e_1 projects to P_1, and $g(v_1)$ takes part in e_1 and projects to variable v_1 of P_1,
then for some e'_2, <Cons[v_1], Diss[v_1]> is true of <e_1, e'_2> relative to π, w, g,

iff for all worlds w in C, perspectival points π, objects d_1, and pairs of eventualities <e_1, e_2>,
if relative to π, w, $e_1 < e_2$, and e_1 projects to P_1, and d_1 takes part in e_1 and projects to variable v_1 of P_1,
then relative to π, w, for some e'_2, d_1 takes part in e'_2 and projects to variable v_1 of P_2, and d_1 is in a less stable position in e'_2 than in e_1.

Informally: if Asterix hits a Roman soldier as shown, then one can find a later event in which Asterix does something less stable.

Note that the cosuppositional requirement on P_1 is weak: it is just that if Asterix hits a Roman soldier as shown, there will be a later event in which Asterix does something that counts as less stable.

By contrast, the cosuppositional requirement imposed on P_2 is more striking.

(127) **Cosuppositional requirement on P_2 in (124)**
(starting with a counterpart of the underlined part of the derivation in (123))
For all worlds w in C, perspectival points π, assignment functions g, and pairs of eventualities $<e_1, e_2>$ such that relative to π, w, $e_1 < e_2$, and e_1 projects to P_1, and $g(v_1)$ takes part in e_1 and projects to variable v_1 of P_1,
if relative to π, w, e_2 projects to P_2, and $g(v_1)$ takes part in e_2 and projects to variable v_1 of P_2,
then relative to π, w, g, $<Cons[v_1], Diss[v_1]>$ is true of $<e_1, e_2>$,

iff for all worlds w in C, perspectival points π, objects d_1, and pairs of eventualities $<e_1, e_2>$ such that $e_1 < e_2$, and e_1 projects to P_1, and d_1 takes part in e_1 and projects to variable v_1 of P_1, and e_2 projects to P_2, and d_1 takes part in e_2 and projects to variable v_1 of P_2, then with respect to π, w, d_1 is in a less stable position in e_2 than in e_1.

Informally: if Asterix hits a Roman soldier as shown and then leaves the room as shown, the Asterix is in a less stable position in the latter than in the former event.

The cosuppositional requirement imposed on P_2 will now be non-trivial and is in fact surprising: despite the violence of the scene in P_1, P_2 is presented as implying something more unstable – this may for instance suggest that there is something deeply unsettling about Asterix's leaving the room; or maybe his earlier misdeed is – at last – coming to haunt his conscience.

Audiovisual examples

The audiovisual examples can be downloaded at the following URL:

https://drive.google.com/file/d/1k4-6296WVOP32LZtBuiihzXgK4S3xHxS

We provide below additional references when they are not given in the main text.

AV00 Heider and Simmel 1944, video cited in:
https://blogs.scientificamerican.com/thoughtful-animal/animating-anthropomorphism-giving-minds-to-geometric-shapes-video/

AV01 Young People's Concert. What Does Music Mean? Written by Leonard Bernstein. Original CBS Television Network Broadcast Date: 18 January 1958.
Retrieved from: https://www.youtube.com/watch?v=9y9fHoB4P2g

AV24, **AV25** Saint-Saëns - Le carnaval des animaux (The Carnival of the Animals) (1886), Conductor: Andrea Licata Royal Philharmonic Orchestra.
Retrieved from: https://www.youtube.com/watch?v=5LOFhsksAYw

AV27 *Psycho*, 1960. Film directed and produced by Alfred Hitchcock, and written by Joseph Stefano. Music by Bernard Herrmann.
Retrieved from: https://www.youtube.com/watch?v=MJdYhJsQF5g

AV31 Strauss, Don Quixote, Variation II. New York Philharmonic Orchestra (conductor: Leonard Bernstein), Carnegie Hall, November 14, 1963.
Retrieved from: https://www.youtube.com/watch?v=7bOVpiuT2kk

AV43, AV44 Strauss, Don Quixote, Variation II. Berliner Philharmoniker (conductor: Herbert von Karajan; cello: Mstislav Rostropovich).
Retrieved from: https://www.youtube.com/watch?v=_6P1WHXKAlk (Gennaro Lettieri)

AV48 Album Artur Rubinstein Plays Chopin Licensed to YouTube by SME (on behalf of RCA Classics).
Retrieved from: https://www.youtube.com/watch?v=P7LDXdjfO5o on December 13, 2019.

AV49 Retrieved from https://www.youtube.com/watch?v=lIIgGWkBx7o on December 13, 2019. The site gives the information: Youra Guller (1895-1981), recorded 1956.

AV50 Les Sylphides. Arrangement for Orchestra by Benjamin Britten, Mono Version. Joseph Levine, American Ballet Theatre Orchestra 1 January 1954.
Retrieved from https://www.youtube.com/watch?v=DQLde6QJXvM on December 13, 2019.

AV51 Les Sylphides: Mazurka, Op. 33 No. 2, arranged by Gordon Jacob. Covent Garden Orchestra, Hugo Ringold. From: Chopin, Glazunov: Les Sylphides (Arranged by Gordon Jacob, Mono Version).
Retrieved from: https://www.youtube.com/watch?v=rGi_mW_elkA

AV52 Chopin, Mazurka in D major Op. 33 No. 2, orchestration Keller 1908. Performed by the Bolshoi Theatre Orchestra conducted by Algis Žiūraitis. Originally on HMV LP ASD 2925.
Retrieved from: https://www.youtube.com/watch?v=6XXHgNfnxoI

AV 54 Heinz Fricke Album Delibes, L.: Coppelia Ballet Suite / Chopin, F.: Les Sylphides (orchestration R. Douglas). Berlin Radio Symphony, Fricke.
Retrieved from https://www.youtube.com/watch?v=jSQbDXEGNUA on December 13, 2019.

AV57 Movement on Chopin’s Mazurka Op. 33 n°2, Performance from 1984 at the American Ballet Theatre, on an orchestration close to Britten’s version.
Retreived from: https://www.youtube.com/watch?v=LBJNc3h7Hp8&t=10m46s

AV62 Disney's The Sorcerer's Apprentice, version from 1940. The Philadelphia Orchestra. Leopold Stokowski.

AV68 Stanley Kubrick, 2001: A Space Odyssey. Warner Bros. Pictures

AV69 Tieu, Lyn; Schlenker, Philippe; Chemla, Emmanuel: 2019, Linguistic Inferences Without Words. *Proceedings of the National Academy of Sciences* 116 (20) 9796-9801.

AV73, AV74, AV75 Bill Mowbray, Uke and Whistle. From: Kitsch In Sync.

AV76, AV77 iMovie audio library, Vintage news short.

AV78, AV79 iMovie audio library, Suspense accents 07.

AV80 Verdi, Simon Boccanegra, Teatro La Fenice. 2014-2015 (conductor: Myung-Whun Chung), RAI.
Retrieved from: https://www.youtube.com/watch?v=H3ChzGDI-AQ

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: 2019, Possible worlds semantics for pictures. In Daniel Gutzman, Lisa Matthewson, Cecile Meier, Hotze Rullmann and Thomas Ede Zimmermann (eds.), *Companion to Semantics*, Wiley.

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.

Blumstein, Daniel T., Bryant, Gregory A. and Kaye, Peter: 2012, The sound of arousal in music is context-dependent. *Biol. Lett.* 8, 744-747

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

Clarke, Eric: 2001, Meaning and the specification of motion in music. *Musicae Scientiae* 5: 213–34.

Clarke, Eric: 2005, *Ways of Listening: An Ecological Approach to the Perception of Musical Meaning*. Oxford University 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.

Cross, I. and Woodruff, G. E.: 2008, Music as a communicative medium. In Botha, R. and Knight, C. (Eds.), *The Prehistory of Language*, Vol. 1, pp. 113–144.

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.

Eitan, Zohar, and Roni Y. Granot: 2006, How music moves. *Music Perception* 23, 3:221-247.

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

Fintel, Kai von 2008. What is presupposition accommodation, again? *Philosophical Perspectives* 22 (1): 137–170.

Gabrielsson, Alf and Lindström, Eric: 2010, The role of structure in the musical expression of emotions. In: *Handbook of Music and Emotion: Theory, Research, and Applications*, eds Juslin P. N., Sloboda J. A., Oxford: Oxford University Press, 367–400

Godoy, R. I. and Leman, M. (Eds.): 2010, *Musical gestures: Sound, movement, and meaning*. Routledge.

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: 2019a, Semantics of Pictorial Space. Manuscript, UCLA.

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

Greenberg, Gabriel: 2021, The Iconic-Symbolic Spectrum. Manuscript, UCLA.

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.

Ilie, G. and Thompson,W. F.: 2006, A comparison of acoustic cues in music and speech for three dimensions of affect. *Music Perception*, 23, 319–29.

Juslin P, Laukka P.: 2003, Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin*; 129(5):770–814.

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

Larson, Steve: 2012, *Musical Forces: Motion, Metaphor, and Meaning in Music*. Indiana University Press.

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.

Maier, Emar and Bimpikou, Sofia: 2019, Shifting perspectives in pictorial narratives. In: M.Teresa Espinal et al. (eds.) *Proceedings of Sinn und Bedeutung 23*, vol. 2, pp. 91–1051. https://semanticsarchive.net/Archive/Tg3ZGI2M/Proceedings23.html

McAdams, Stephen: 1984, *Spectral Fusion, Spectral Parsing, and the Formation of Auditory Images*. Unpublished PhD thesis, Stanford University. https://ccrma.stanford.edu/files/papers/stanm22.pdf

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.

Parsons, Terence: 1990, *Events in the Semantics of English*. MIT Press.

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

Pasternak, Robert and Tieu, Lyn: 2020, Co-linguistic content projection: From gestures to sound effects and emoji. Manuscript, ZAS Berlin and U. of Western Sydney.

Patel-Grosz, Pritty; Grosz, Patrick Georg; Kelkar, Tejaswinee & Jensenius, Alexander Refsum (2018). Coreference and disjoint reference in the semantics of narrative dance, In Uli Sauerland & Stephanie Solt (ed.), *Proceedings of Sinn und Bedeutung* 22, vol. 2, ZASPiL 61. Leibniz-Zentrum Allgemeine Sprachwissenschaft (ZAS). Chapter in Vol. 2. s 199 - 216

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

Roberts, Craige: 2012, Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics & Pragmatics*, 5 https://semprag.org/article/view/sp.5.6

Rodriguez, Hugo: 2021, *Sémantique et pragmatique de la musique. Une approche cognitive basée sur le travail de Philippe Schlenker et sur les œuvres de Franz Liszt*. Doctoral dissertation, Université

Libre de Bruxelles.

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: 2012, Maximize Presupposition and Gricean reasoning. *Natural Language Semantics* 20, 4: 391-429

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: 2021, Iconic Presuppositions. *Natural Language & Linguistic Theory* 39:215–289.

Sievers, B., Polansky, L., Casey, M., & Wheatley, T.: 2013, Music and movement share a dynamic structure that supports universal expressions of emotion. Proceedings of the National Academy of Sciences, 110, 70-75. doi:10.1073/pnas.1209023110

Sievers B, Lee C, Haslett W, Wheatley T. 2019 A multi-sensory code for emotional arousal. *Proc. R. Soc. B* 286: 20190513. http://dx.doi.org/10.1098/rspb.2019.0513

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

Stalnaker, R. 2002. Common ground. *Linguistics and Philosophy* 25 (5–6): 701–721.

Steedman, Mark: 2002, Helmholtz' and Longuet-Higgins' Theories of Consonance and Harmony. Unpublished Tutorial Paper.

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: 2018, 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

Zaradzki, Léo: 2021, *Les évènements en sémantique linguistique et musicale*. Doctoral dissertation, University of Paris-Diderot.