Linguistic inferences from pro-speech music

Musical gestures generate scalar implicatures, presuppositions, supplements, and homogeneity inferences

Léo Migotti; Janek Guerrini

April 2022

Abstract

Language has a rich typology of inferential types. It was recently shown that subjects are able to divide the informational content of new visual stimuli among the various slots of the inferential typology: when gestures or visual animations are used in lieu of specific words in a sentence, they can trigger the very same inferential types as language alone (Tieu *et al.*, 2019). How general are the relevant triggering algorithms? We show that they extend to the auditory modality and to music cognition. We tested whether pro-speech musical gestures, i.e. musical excerpts that replace words in sentences, can give rise to the same inferences. We show that it is possible to replicate the same typology of inferences using pro-speech music. Minimal and complex musical excerpts can behave just like language, gestures, and visual animations with respect to the logical behavior of their content when embedded in sentences. Specifically, we found that pro-speech music can generate scalar implicatures, presuppositions, supplements, and homogeneity inferences.

Contents

1 Introduction

Natural language semantics has long evidenced the existence of a rich typology of content, showing that the meaning of sentences can be split up into different content types (Schlenker, 2016; Kadmon, 2001). Let us take the example of presuppositions.[1] Sentence (1)a conveys two kinds of information: the speaker asserts that John does not smoke now, while taking for granted that John used to smoke. Such inferences, presuppositions, survive in different logical environments. When turning the sentence into a question in (1)b, the assertion that John stopped smoking is now questioned, but the presupposition that he used to smoke remains.

(1) a. John stopped smoking.
$\rightsquigarrow$ John used to smoke.

b. Did John stop smoking?
$\rightsquigarrow$ John used to smoke.

It has been argued that such a projection pattern (namely, here, the fact that the presupposition is preserved under question formation[2]) is lexically encoded, i.e. determined on the basis of word meaning (Levinson, 2000; Abusch, 2002). The meaning of "stop" would for instance be divided as in (2).

(2) *x* stops *Q*-ing.
- At-issue: *x* does not *Q*.
- Presupposed: *x Q*-ed before.

Alternatively, some theorists have claimed that general algorithms can predict when an inference triggered by a given word is treated as a presupposition, in part because across languages the same projection patterns are observed (Simons *et al.*, 2010; Tonhauser *et al.*, 2013; Abusch, 2010; Abrusán, 2010; Schlenker, 2021). On this view, humans can divide asserted content from presupposed content productively and on the fly. Here, we propose to test the productivity of these mechanisms by making use of non-linguistic items such as music: if such general and productive procedures exist, then the same inferential patterns should arise from non-linguistic stimuli. In other words, we infer the existence of a hardwired, general-purpose algorithm that subdivides meaning into at-issue and non-at-issue from people's ability to endorse inferences containing both types of content on the fly. This can be conceived of as a variant of the argument of the poverty of the stimulus (PoS) (Chomsky, 1980, 1988). PoS states that the sentences children are exposed to when learning a language do not contain enough of the information needed to develop a thorough understanding of the grammar of the language. Hence, learning is at least partly made possible by a core of innate linguistic mechanisms.

Similarly for our case: contextual conditions such as knowledge of specific words of a given language cannot be responsible for systematic inferential behavior in presence of stimuli subjects have plausibly never been exposed to. Consequently, there is no real alternative to the conclusion that humans have an algorithm available to adequately arrange content in at-issue and non-at-issue before experience with a given linguistic item. In both cases, systematic behavior (competent use of a language for the general PoS, systematic inferential behavior in our case) is not backed by enough exposure to justify its learning, and thus can only be caused by subject-internal structures (Universal Grammar for the general PoS, general triggering algorithms in our case).

Both theoretical and experimental research on pro-speech gestures and pro-speech visual animations (i.e. gestures or visual animations replacing words in sentences) have corroborated these predictions. Schlenker (2019a) developed a typology of embedded gestural depictions replacing, following, or co-occurring with words, showing that, in particular, the content of pro-speech gestures can be divided among familiar slots of the inferential typology. Tieu *et al.* showed this experimentally. Consider (3) from Tieu *et al.* (2019):

(3) a. The student REMOVE_GLASSES.
$\rightsquigarrow$ The student currently has glasses on.

[1] Here we use the example of presuppositions as an illustration of the main argument of the paper. The argument however applies similarly to the three other inferences tested: scalar implicatures, supplements and homogeneity inferences. A summary of the different inferential mechanisms can be found in Appendix III.

[2] Generally, the projection *problem* refers to the computation of the presupposition and asserted content of a sentence from the presupposition and asserted content of its constituents. By contrast, projection *tests*, such as the family-of-sentences test (so-called in Chierchia & McConnell-Ginet (2000)), i.e. embedding under negation, modality or question, are used to check whether a given proposition is a presupposition (Heim, 1990, 1992; Stalnaker, 1974; Geurts, 1999; Beaver, 2001; Chemla, 2009)

b. Will the student REMOVE_GLASSES?
⇝ The student currently has glasses on.

Here, REMOVE_GLASSES stands for the gesture mimicking someone removing their glasses. Sentence (3)a presupposes that, at the time of the utterance, the student has glasses on. Similarly, the question in (3)b still presupposes that the student has glasses on, just like the presupposition that John used to smoke was preserved under question in (1). Subjects were shown to be able to access this presuppositional content significantly more than control inferences. In general, Tieu *et al.* found that embedded visual stimuli can give rise to the same inferential types as purely linguistic ones. This suggests that the algorithm humans apply to divide at-issue content from non-at-issue content must be productive, and, because the stimuli were visual, it cannot be limited to language narrowly intended.

How general is the algorithm that divides at-issue content from non-at-issue content in the case of presuppositions, and how general are other inferential algorithms producing, for instance, scalar implicatures, supplements, or homogeneity inferences? Here we ask whether they extend to the auditory modality and to music cognition. We used pro-speech music, i.e. musical excerpts that replace words in sentences, henceforth *musical gestures*.[3] Building our paradigm after Tieu *et al.* (2019), we tested whether musical gestures can give rise to the same inferential typology as language, pro-speech gestures and pro-speech visual animations. For example, in a context involving a person that was hiking, sentence (4) behaved just like a change-of-state verb like "stop", and triggered the presupposition that the hiker is down a mountain both when embedded in a declarative sentence as in (4)a, and when it is embedded in an interrogative sentence as in (4)b.[4]

(4) a. The hiker will UPWARD_SCALE.[5]
⇝ The hiker is at the foot of a mountain.

b. Will the hiker UPWARD_SCALE?
⇝ The hiker is at the foot of a mountain.

UPWARD_SCALE stands for a minimal musical stimulus of a classical C major scale played by a harp. Each note of the scale was played one after the other following a standard rise in frequency at a 990 notes/s rate to approach the sound of a real harp *glissando*, often used in classical music. In general, because of the intuitive mapping between gesture and meaning (Schlenker, 2019a), the informational content of iconic gestures can be grasped even in absence of previous exposure (Schlenker, 2017). The reader has plausibly never been exposed to (4)a or (4)b, but the content of the musical gesture UPWARD_SCALE can still be productively divided between at-issue content, *viz.* the hiker is going up, and presuppositional content, *viz.*at the time of utterance, the hiker is located at a low point in space.[6] Crucially, the inference goes through even though (4)b is a question, which is a classical test for presupposition projection. Besides presuppositions, we try to replicate the typology of inferences already replicated with gestures and visual animations in Tieu *et al.* (2019), including scalar implicatures, supplements, and homogeneity inferences.

Here we are merely interested in the semantic and pragmatic behavior of the informational content of musical sounds. This paper does not aim at deciding between different theories of the respective inferences we tested; rather, it reports experimental data that any theory should account for.[7] Still, in some cases these results make it possible to exclude a significant class of theories, as for instance in the case, mentioned above, of lexical theories of presuppositions.

[3] Although 'pro-speech music' [literally, music replacing words] is a specific kind of musical gestures [i.e. the iconic musical motives or excerpts roughly used in lieu of gestures in Tieu *et al.* (2019)], we use 'pro-speech music' and 'musical gestures' interchangeably throughout this paper.

[4] Except for the paradigm testing supplements in section 3.3, we mainly used basic scales, drum sounds or isolated tones. Our definition of these as music could be contested because of their simple nature. However, even if these stimuli did not count as music, our claims on the generality of the algorithm that divides content in at-issue and non-at-issue would remain unaltered.

[5] All musical gestures can be directly accessed by clicking on the hyperlinks.

[6] As pointed out by an anonymous reviewer, we cannot know when precisely participants drew the inference. However, whenever the inference is actually triggered, it is unclear how our predictions would be different at this stage. As the same issue was present in Tieu *et al.* (2019), we assume that our predictions would not have been different.

[7] For instance, our data does not allow us to decide between a grammatical or a neo-Gricean theory of scalar implicatures, but allows us to claim that regardless of the details of the algorithm responsible for scalar implicatures, this algorithm extends to music cognition and must therefore be domain-general.

Neither is the paper about how the content of pro-speech music relates to the actual musical properties of the stimuli, i.e. about musical meaning and how it is derived. Still, we provide first insights into how this meaning can interact with the logical structures of language, which suggests the existence of a non-trivial informational content in music. We leave open the challenging question of whether music triggers more sophisticated content, especially outside of a linguistic context.

Before we move on, let us address a potential worry concerning the methodology used here and in Tieu *et al.* (2019). There is the possibility that pro-speech music is systematically translated into words. In this case, there would be the possibility that the relevant inferences arise because they are lexically encoded in the words of the translation. Consequently, our results would be uninformative about the generality of the algorithm dividing content in at-issue and non-at-issue. Two reasons militate against this hypothesis. First, just as in Tieu *et al.* (2019), subjects of our experiment were able to interpret fine-grained gestural iconic information that was absent from the words of the closest verbal translation. Second, such verbal translations would make iconic dimensions at-issue when not encoded in a verbal translation that lexically makes them non-at-issue, a behavior that seems to be excluded by the logical tests we provide in section 4.

This paper is structured as follows. In section 2, we detail the experimental setup. In section 3.1 we present the paradigm we used to test for scalar implicatures. Two stimuli involving respectively one and three repetitions of a drum sound competed and gave rise to scalar implicatures. In section 3.2, we present the paradigm we used to test for presupposition projection, already introduced in (4). While the presupposition was preserved under question formation, participants did not behave as expected in the classical test under "none". We discuss in detail possible explanations of the difference between our results and Tieu *et al.*'s. We then move on to other linguistic inferences and introduce a paradigm testing for supplements in section 3.3. Pro-speech music did indeed behave like a supplement, yielding the typical conditional projection. In section 3.4, we test homogeneity inferences from pro-speech music. There are different theoretical accounts of such inferences: some attribute the existence of homogeneity inferences to the noun phrase, others to the predicate. We prove "by case" that pro-speech music gives rise to homogeneity inferences both when it replaces the noun phrase and when it replaces the predicate. Finally, in section 4, we discuss in detail why it is unlikely that pro-speech music is systematically translated into words.

2 Methods

Stimuli were recordings of French sentences with musical gestures (either specifically generated for our purposes or pre-existing) either artificially generated through GarageBand (version 10.3.5) or taken from real music replacing one or some words.

We collected data on 68 participants. We excluded non-native French speakers, participants who had already taken one of our pilot experiment and participants who had failed the two attention checks, which left us with 53 subjects. Participants were recruited through the French platform Crowdpanel, and online informed consent was obtained for each of them. After a short training on three examples of hybrid sentences containing music, and for each stimulus, each participant had to assess to what extent the presented inference followed from the auditory stimulus (i.e. the spoken sentences with embedded music), using a slider bar ranging from 0, labeled *totally disagree* (in French: *pas du tout d'accord*), to 100, labeled *totally agree* (in French: *tout à fait d'accord*) (cf. Figure 1)[8]. Stimuli played automatically for each trial, but participants could listen to the stimulus as many times as needed. The experiment was set up on Qualtrics. All stimuli were fully randomized. In each section, we report comparisons of generalized mixed-effects models of the data using R (version 3.6.2.) to assess the contribution of each factor of interest to the model (R core Team (2016); Barr *et al.* (2013)). The details, justifications, simplifying procedures and coefficients of the models are available in Appendix II. The details of the stimuli are given in the next sections, for each type of inference.[9]

[8] As pointed out by a reviewer, it cannot be ruled out that when asked about inferences, subjects understand that they must guess the right word in the stimuli, e.g. 'climb'. Indeed, the target sentence was visible while listening to the stimulus, just like in (Tieu *et al.*, 2019). Since this issue applies to all of this literature, we leave it as a problem for future research to understand if and how participants consider the lexical material from the target sentence to be relevant to the comprehension of the stimulus.

[9] All material including stimuli, design files and analysis scripts are accessible at https://osf.io/hw45u/?view_only=89f983db777f49e9a6f5b41b3dea60d6. Material was uploaded prior to the beginning of the data collection - See PREREGISTRATION folder. Final results and statistics are available in the RESULTS folder.

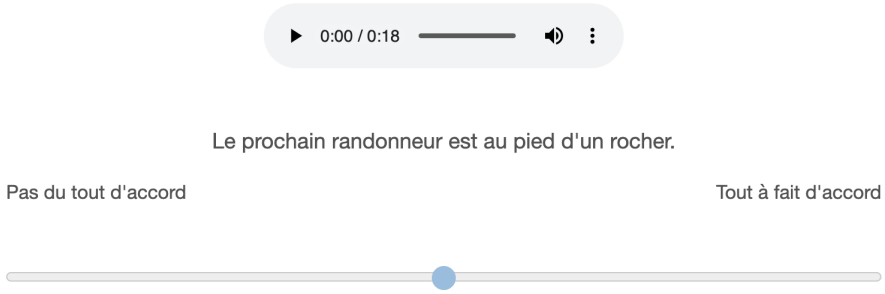

Figure 1: An example of a stimulus discussed in 3.2

3 Replication of the typology of inferences

3.1 Scalar implicatures

3.1.1 In language and from visual stimuli

Scalar implicatures convey an implicit, strengthened meaning beyond the explicit or literal meaning of an utterance. Originally, they were viewed as arising from pragmatic reasoning on the speaker's communicative intentions (Grice, 1975). According to neo-Gricean approaches (Horn, 1972; Sauerland, 2004; van Rooij & Schulz, 2004; Chierchia *et al.*, 2012), if a sentence S evokes S' and S' is more informative than S (or not less informative than S, depending on the specific approach), speakers infer that S' is false. For instance, in (5)a the speaker chose to utter "some" instead of the logically stronger "all", and is therefore inferred to mean *some but not all* (Sauerland, 2012).

More recently,some theorists have argued that the mechanism responsible for scalar implicatures is instead a grammatical one (Chierchia *et al.*, 2012). On this view, a silent operator with a meaning very similar to *only* is applied to scalar sentences: the inference in (5)c is the result of interpreting "some" in (5)a as *only some*, as shown in (5)b.

(5) a. He read some of the books.
b. *Exh* (He read some of the books.)
⇝ c. He read some, but not all of the books.

Either theory has at its core competition among alternatives interacting with informativity considerations.[10] Scalar implicatures were recently shown to arise between alternatives realized by means of a gesture or a visual animation provided in a context (Tieu *et al.*, 2019). For instance, in the positive environment in (6) where TURN_WHEEL stands for the gesture mimicking a driver turning a wheel, we understand that TURN_WHEEL does not only mean *to turn* but *to turn somewhat, but not a lot*. TURN_WHEEL competes with the more informative gesture TURN_WHEEL_COMPLETELY, which is not used, thus taken to be false.

(6) a. He will TURN_WHEEL_COMPLETELY.
⇝ He will turn the wheel completely.

b. He will TURN_WHEEL.
⇝ He will turn the wheel, but not completely.

The inference in (6)b. could be explained by the fact that TURN_WHEEL semantically means *to turn somewhat, but not completely*. In this case, however, we would expect *not*-TURN_WHEEL to mean *to not turn somewhat*.

[10]For presentational clarity, we choose to present scalar implicatures by going through the Gricean reasoning. We then mention the alternative theory, the grammatical account of scalar implicatures. In any case, these should be viewed as placeholders for any theory of scalar implicatures. Note that any theory of scalar implicatures predicts that if alternatives are provided in the context, implicatures should be derived. This is, in a way, a sanity check to confirm that the inferential mechanisms work as expected with pro-speech sounds, and that there can be competition among musical excerpts.

(7) a. He will not TURN_WHEEL_COMPLETELY.
⇝ He will turn the wheel, but not completely.

b. He will not TURN_WHEEL.
⇝ He will not turn the wheel at all.

We rather understand from (7)b that the wheel was not turned at all. The informativity pattern gets reversed under negation, just like in language. *not*-TURN_WHEEL is now logically stronger than not-TURN_WHEEL_COMPLETELY. The inferences in (7) show that there is competition between the two gestures.

3.1.2 From music

We reasoned that realizations with different numbers of repetitions of a sound could be represented by speakers as a logical scale (Horn, 1972). We generated two different realizations of the timpani sound DRUM (from GarageBand instruments library): the weaker DRUM×1 and the stronger DRUM×3. We first set up a context introducing the stronger alternative:

(8) **Context (positive environment):** Jean boxes regularly at the gym. During last week's workout, he had a lot of energy, and was able to DRUM×3.

In the target premise (9), DRUM×1 evokes the more informative alternative DRUM×3, introduced in the context in (8):

(9) **Target premise**
This week, Jean will DRUM×1.

Original stimulus

Cette semaine, Jean va DRUM×*1.*
This week, Jean will DRUM×1.

If DRUM×1 forms a scale with DRUM×3, the stronger alternative DRUM×3 should be taken to be false if not realized. We expect speakers to draw the inference that John will box somewhat, but not a lot:

(10) **Target inference**
This week, Jean is going to box somewhat, but not a lot.

In the control stimulus in (11), we did not vary the number of repetitions:

(11) **Control premise**
This week, Jean will DRUM×3

Original stimulus

Cette semaine, Jean va DRUM×*3.*
This week, Jean will DRUM×3.

Because in (11) the informativity of the stimulus is not manipulated, we expect speakers to stick with the context and endorse the baseline inference in (12) when given the control premise.

(12) **Baseline inference**
This week, Jean will box a lot.

Just like for linguistic stimuli, the critical test is negation. Without such a test, the expected inference from the positive environment in (12) may be explained otherwise. Namely, DRUM×3 may convey that Jean will box a lot simply because it means *to punch something exactly three times*, and not because it constitutes a logically stronger alternative than DRUM×1. To rule out this alternative explanation, we set up a context in which both alternatives were introduced:

(13) **Context (negative environment):** Jeanne is boxing at the gym. At last week's session, she had a lot of energy, and was able to DRUM×3. But during the second week of training, she did not DRUM×1.

If the two realizations form a scale, under negation, informativity should be reversed: *not*-DRUM×1 constitutes a more informative alternative with respect to *not*-DRUM×3:

(14) **Target premise**
This week, Jeanne will not DRUM×3.

Original stimulus

Cette	*semaine,*	*Jeanne*	*ne*	*va*	*pas*	DRUM×*3.*
This	week,	Jeanne	NEG	will	not	DRUM×3.

If DRUM×3 means *to punch something exactly three times*, we expect its negation to mean *not punch something exactly three times*, i.e. two or less, or four or more. If on the other hand the two realizations form a scale, as we argue, we expect *not*-DRUM×3 to convey that although not a lot, some boxing still occurred, as in (15) below.

(15) **Target inference**
This week, Jeanne will box somewhat, but not a lot.

And similarly for the weaker realization: if DRUM×1 means *to punch something exactly once*, then *not*-DRUM×1 should mean *not punching something exactly once*. This yields an inference pattern different from what is expected under our hypothesis that *not*-DRUM×3 forms a scale with *not*-DRUM×1. In this case, *not*-DRUM×1 should convey that there was no boxing at all.

(16) **Control premise**
This week, Jeanne will not DRUM×1.

Original stimulus

Cette	*semaine,*	*Jeanne*	*ne*	*va*	*pas*	DRUM×*1.*
This	week,	Jeanne	NEG	will	not	DRUM×1.

We thus expect *not*-DRUM×1 to be interpreted as *did not box at all* rather than *did not punch something exactly once*, which would allow for boxing more than once:

(17) **Baseline inference**
This week, Jeanne will not box at all.

3.1.3 Results

A scalar implicature triggered by the target premise would lead participants to endorse the target inference more than the baseline inference. Such a difference could be due to an *a priori* preference for the target inference, independently of the target premise. For that reason, we looked at the interaction between target *vs* control premise and target *vs* baseline inference. We did find a significant interaction between the two factors that rule out this possibility in both environments ($\chi^2 = 170$, $p < 0.001$ in the positive environment; $\chi^2 = 21$, $p < 0.001$ in the negative environment), compatible with the triggering of a scalar implicature resulting from the competition of two musical alternatives, namely DRUM×1 and DRUM×3.

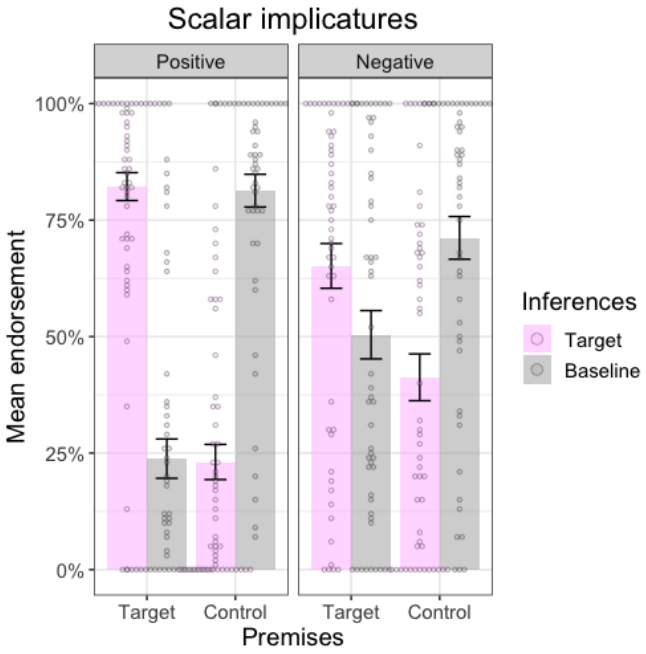

Figure 2: Mean endorsement rate for scalar implicatures

3.1.4 Discussion

In the positive environment, as expected, DRUM×1 competes with the more informative alternative DRUM×3 so that DRUM×1 is understood as *to box somewhat but not a lot.* In the negative environment, *not*-DRUM×3 competes with *not*-DRUM×1. Thus *not*-DRUM×3 was interpreted as *box somewhat, but not a lot*. This shows that when two realizations have different logical informativity, they compete with each other just like alternatives compete in the case of gestures and visual animations in Tieu *et al.* (2019), leading to the generation of scalar implicatures by means of non-purely linguistic means.[11]

3.2 Presuppositions

3.2.1 In language and from visual stimuli

A presupposition is a meaning component that is taken for granted that must follow from the local context of an expression, and which sentences inherit across different logical operators, such as negation (Stalnaker, 1974; Heim & Kratzer, 1998). In language, some particular expressions are associated with the triggering of presuppositions. For instance, in (18), the presupposition that John used to smoke is triggered by "stop". In general, change-of-state verbs (e.g. "start", "become", "stop") all presuppose their initial state (Abrusán, 2010). For John to be able to stop smoking, John must have been smoking to begin with.

Presuppositional content is untouched by logical operations that change the meaning of a sentence, such as negation, modality, and questions (Abusch, 2010; Abrusán, 2011; Chemla, 2009). For instance, the question in (18)b still presupposes that John used to smoke.

[11] In our stimuli, alternatives were provided in the context: the alternative, e.g. DRUM×3, to the musical gesture, DRUM×1, was salient in the context. For this reason, our results don't speak to the issue of alternative generation. Our point is that scalar implicatures can be triggered by pro-speech music as long as two alternatives are available. In other words, whatever the mechanism responsible for alternative generation, subjects are able to interpret one musical alternative as logically stronger than the other, have it compete with the other, and finally draw a scalar implicature from this process. However, if scalar implicatures were in fact to arise when alternatives are not directly given, then a general theory of alternatives could be needed. This is an exciting question for future research. Schlenker (2020) discussed such a theory in relation to Katzir (2007) in the case of gestures. As to music, it would not be surprising that implicatures are triggered whenever a musical gesture competes with a contextually salient (but not explicitly given) more informative alternative.

(18) a. John stopped smoking.
⇝ John used to smoke.

b. Did John stop smoking?
⇝ John used to smoke.

Tieu *et al.* (2019) found that presuppositions could also be generated by means of pro-speech gestures and visual animations. First, they were preserved under question formation:

(19) **Context** 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 an architecture firm. Jane is watching the union representatives and says:

(20) "Will the union representatives' antennae GREEN_TO_BLUE?" [12]
(animation content: **bar is green at first, then slowly whole bar goes blue**)

⇝ The union representative is currently not in a meditative state.
↛ The union representative is currently in a meditative state.

The context used in (19) introduces the two states between which the transition denoted by the visual animation GREEN_TO_BLUE can be made. When turning from green to blue, the embedded visual animation triggers the presupposition that the initial state denoted by the green color holds before the change of state takes place, i.e. the individual is not in a meditative state at the moment of utterance.

Evidence for presuppositions being triggered by visual animations was found under "none", too, as shown in (21) below.

(21) "None of the union representatives' antennae will GREEN_TO_BLUE"

⇝ None of the union representatives are currently in a meditative state.
↛ Some of the union representatives are currently in a meditative state.

Tieu *et al.*'s (2019) choice to test for presupposition under "none" instead of under negation was motivated by Chemla (2009) results on strong universal inferences under "none" in French. This choice has significant theoretical advantages. Some apparent presuppositions that project out of negative environments such as (22)a. may be due to a scalar implicature and may not be strong evidence for presuppositional content. In (22) TURN_WHEEL_COMPLETELY stands for the gesture mimicking a driver turning a wheel completely. The fact that the agent is in front of a steering wheel may come as a consequence of the scalar implicature that although there was no complete turning of the wheel, there was still some turning (Tieu *et al.*, 2019). For there to be some turning, the agent must be in front of a wheel in the first place.

(22) He did not TURN_WHEEL_COMPLETELY.

a. The agent is in front of a steering wheel.

Under "none", however, scalar implicatures project existentially, not universally, as shown in (23) below (Chemla, 2009).

(23) No student read all the books.

⇝ (At least) one student read (at least) some of the books..

The universal projection under 'none' found by Tieu *et al.* (2019) and reported in (21) thus constituted very strong evidence in favor of presupposition projection.

[12] The original sequence of sentences and the visual animations can be accessed at Tieu et al.'s supplementary materials page: https://mfr.au-1.osf.io/render?url=https://osf.io/v5xa3/?direct%26mode=render%26action=download%26mode=render

3.2.2 From music

We tested whether pro-speech music could generate presuppositions. Consider the context below:

(24) ***Context***: Some hikers are hiking in the mountains, where there can be significant drops and peaks in elevation. They alternate between reaching the top and the foot of steep rocks. Two of them, who finished first, are talking while waiting for the others.

We used an upward scale played by a harp to evoke a hiker going up a mountain, as in (25).

(25) **Target premise (question)**
One asks the other: "Will the next hiker UPWARD_SCALE?"

Original stimulus:

L'	*un*	*demande*	*à*	*l'*	*autre*	*:*	*"Est*	*ce*	*que*	*le*	*prochain*	*randonneur*	*va*	UPWARD_SCALE*?"*
The	one	asks	to	the	other	:	"Is	it	that	the	next	hiker	will	UPWARD_SCALE?"

Question formation is a classical test for presupposition (Chemla, 2009; Tieu *et al.*, 2019). In (25), if the initial state of the musical scale is presupposed rather than at-issue, the sentence should presuppose that the hiker is at the bottom of a rock. We thus expect the target inference in (26) to be more endorsed than its negation, the baseline inference in (27).

(26) **Target inference**
The next hiker is at the foot of a rock.

(27) **Baseline inference**
The next hiker is not at the foot of a rock.

Tieu *et al.* (2019) chose to test universal projection under "none", and we follow suit to have a clear point of comparison between acoustic and visual gestures.

(28) **Target premise under "None"**
One tells the other: "None of the hikers will UPWARD_SCALE".

Original stimulus

L'	*un*	*dit*	*à*	*l'*	*autre*	*:*	*"Aucun*	*des*	*randonneurs*	*ne*	*va*	UPWARD_SCALE.
The	one	tells	to	the	other	:	"None	the	hikers	NEG	will	UPWARD_SCALE.

Given Tieu *et al.*'s 2019 results, we thus expected the universal inference in (29) to be more endorsed than its negation, the baseline inference in (30).

(29) **Target inference**
Each of the hikers is at the foot of a rock.

(30) **Baseline inference**.
Not all of the hikers are at the foot of a rock.

3.2.3 Results

As explained in Appendix II, here we expected a significant effect of inference type on the endorsement of the premise: the inference containing the presupposition was expected to be significantly more endorsed than the baseline inference negating the presupposition. Under question formation, as expected, we found a significant effect of inference type ($\chi^2 = 6$, $p < 0.05$), compatible with the triggering of a presupposition. We also found a significant effect of inference type in the 'none' environment, ($\chi^2 = 22$, $p < 0.001$), where, a bit more surprisingly, the baseline inference that did not contain the presupposition was more highly endorsed than its counterpart containing the projected presupposition. We discuss some hypotheses regarding this pattern in the following subsection.

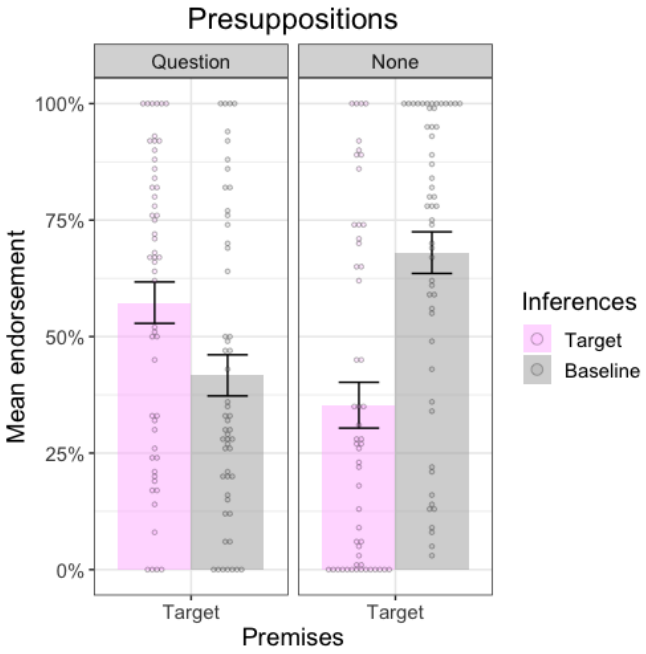

Figure 3: Mean endorsement rate for presuppositions

3.2.4 Discussion

In the interrogative environment, as expected, participants judged the initial state of UPWARD_SCALE to be true at the time of the utterance. Participants computed the initial state as non-at-issue, a behavior that matches the inferential pattern of change-of-state verbs. To wit, if the initial state had been at-issue, the question would have targeted the information that the hiker is at the bottom, amounting to something like "Is the hiker at the bottom and will he go up?". Instead, the question only targeted the change of state, asking something like "Will the hiker go up?". This supports our hypothesis that there is a productive algorithm at work in these cases, as subjects had never been exposed to such a sentence.

By contrast, UPWARD_SCALE embedded under "none" did not give rise to the expected universal inference that "Each of the hikers was at the foot of a rock". We discuss several reasons for this observed pattern below.

First, it is worth mentioning that the data on projection under "none" are in part disputed: initially, Heim (1988) predicted universal projection, while Beaver (1994) predicted existential projection. Recent approaches find a complex picture (Zehr *et al.*, 2016). We briefly review the theoretical insights and empirical evidence on this topic from Chemla (2009). Consider sentence (31), but see Zehr *et al.* (2016) for a complete review.

(31) None of the players stopped smoking.

a. At least one player used to smoke and none of them stopped smoking.

b. All of the players used to smoke and none of them stopped smoking.

c. None of the players both used to smoke and stopped smoking.

d. None of the players (who smoke) stopped smokin. $\vDash$ All of the players who smoke smoke. $\vDash$ At least one player smokes and none of them stopped smoking.

Although all theories of presupposition can derive more or less straightforwardly both types of projection, there is disagreement on whether the existential or the universal is the by-default reading. Sentence (31) could project weakly, as in ((31) a.), as predicted by a class of theories that stress existential projection (Beaver, 1994; Van Der Sandt, 1992; Geurts, 1998; Mandelkern, 2016).

Other theories predict that sentences like (31) display a default universal presupposition projection (Heim, 1988; Schlenker, 2008, 2010; George, 2008; Fox, 2013; Chemla & Schlenker, 2012; Mayr & Sauerland, 2016). The presence of existential projections is accommodated by these theories, as any existential can be captured as the result of a (conditional) domain restriction on the "none" - "none of the players (who smoked) stopped smoking", as in (31)d., where the universal, adequately restricted, gives rise to an inference equivalent to an existential.

A third possible reading is a consequence of local accommodation, as in (31)c. Presuppositions can be interpreted within the scope of a logical operator that makes them at issue. In (32)a., the presupposition is interpreted within the scope of negation and the resulting truth conditions end up amounting to the conjunction of the supposed presuppositional content with the supposed at-issue content as in (32)b.

(32) a. John didn't stop smoking, he didn't even start!

b. John didn't begin and stop a smoking habit, he didn't even start!

The empirical evidence is not more uniform. As mentioned above, Chemla (2009) showed the availability of a universal reading in *none* sentences. In similar studies, it was argued that the most robust inference is the existential one (Sudo *et al.*, 2012; Geurts & van Tiel, 2016). All studies found endorsement for both kinds of inferences, and Zehr *et al.* (2016) proposed that both the existential and the universal inference are accessible to subjects, leaving open the question of the typology of the logical behaviors of presuppositions under *none*.

Turning to our result, it seems that a positive result would have been strong evidence of our hypothesis (all universal projections are presuppositions). However, in view of the complexity of the behavior of presupposition triggers under "none", while the lack of universal projection weakens a bit our conclusion, it does not (at all) refute it (not all non-universal projections are non-presuppositions/not all presuppositions project universally).

A second point of discussion is that in spite of all this, the question remains of why we found a result different from Tieu *et al.* (2019). We here give two tentative explanations.

First, consider again our context in (24), reported below in (33):

(33) Some hikers are hiking in the mountains, where there can be significant drops and peaks in elevation. They alternate between reaching the top and the foot of steep rocks. Two of them, who finished first, are talking while waiting for the others.

It is possible that participants inferred that not all of the hikers were at the bottom of a rock because the two characters discussing were at the top of a rock already, given that the context states that *they have already finished.* Let us again look at sentence (28), reported below in (34).

(34) **Target premise (question)** One tells the other: "None of the hikers will UPWARD_SCALE?"

If the domain of "the hikers" in the target inference in (29) *includes* the two hikers discussing, the target inference in (29) should be false. In Tieu *et al.* (2019), instead, it was clear that the speaker and the object matter were distinct.

One further possible explanation for the observed difference is that it might be easier to convey absolute initial states in the visual modality than in the auditory modality. In Tieu *et al.* (2019), the visual animation of a green bar becoming blue (GREEN_TO_BLUE) was used to refer to the color change in the antennae of aliens introduced in the context. In our own paradigm, the first note of the musical scale was used to refer to the level at which the hiker was located in space. While in our context any point on the rock can be denoted by the first note of the scale, the initial state "green" in Tieu et al.'s stimulus can only be interpreted as absolute, i.e. no other point on the color spectrum is compatible with this initial state. Then, some of the hikers may fail to climb up while (i) being in a position to climb up, i.e. not at the top, and (ii) not being at the foot of the rock (but higher). This would explain why subjects thought that not all of the hikers were at the foot of a rock.

3.3 Supplements

3.3.1 In language and from visual stimuli

Supplements are inferences triggered by non-restrictive relative clauses that behave like independent sentences (Potts, 2004; Schlenker, 2019a). For instance, sentence (35)a below is understood as (35)b and not as (35)c.

(35) a. It is unlikely that Robin lifts weights, which is harmful.
b. It is unlikely that Robin lifts weights. Lifting weights is harmful.
c. It is unlikely that Robin lifts weights and that this is harmful.

Tieu *et al.* (2019) found that both gestures and visual animations used in lieu of a non-restrictive relative clause gave rise to the conditional presupposition characteristic of supplements. For instance, in (36), the gesture HIT replaces a non-restrictive relative clause similar to "which involves hitting". The informational content of the gesture is not at-issue here, i.e. it is not target by "if", hence the inference that if the event under question happens, it will involve some hitting.

(36) If June bugs a classmate today - HIT, she will get a detention.
$\rightsquigarrow$ If June bugs a classmate today, it will involve hitting her.

By contrast, in (37), the content of the gesture is made at-issue by the use of "and does so *like this*". The way in which the bugging happens, i.e. by hitting, is now made at-issue and thus targeted by the conditional, hence the weaker inference that hitting will not necessarily be involved.

(37) If June bugs a classmate today and does so like this - HIT, she will get a detention.
$\rightsquigarrow$ If June bugs a classmate today, it will not necessarily involve hitting her.

3.3.2 From music

We used an excerpt of real film music by American composer Bernard Herrmann from the soundtrack of the movie *Psycho* to denote a scary dog coming closer to the character in the context shown in (38).[13] We reasoned that a rich musical gesture with a non-ambiguous emotional content could be informative about the scene while being interpreted as a non-restrictive relative clause.

(38) **Context:** Marie is walking back home. She spots a dog on the other side of the street. She sometimes worries about dogs, because some of them can be vicious - PSYCHO.

In the target premise in (39), PSYCHO behaves like a non-restrictive relative clause. It conveys that if the dog comes closer, it will look somewhat dangerous or scary.

(39) **Target premise**
If the dog comes to her - PSYCHO, Marie will cross the street.

Original stimulus

Si le chien s' approche d' elle - PSYCHO, *Marie changera de trottoir.*
If the dog self approaches to her - PSYCHO, Marie changes of sidewalk.

If the musical gesture PSYCHO logically behaves like a non-restrictive relative clause, we expect the target sentence in (39) to give rise to the target inference in (40), in which the behavior of the dog is not at-issue, thus not targeted by the conditional.

(40) **Target inference**
If the dog comes to Mary, it will look vicious.

To ensure that the musical gesture PSYCHO was not interpreted as a non-restrictive relative clause in the control premise in (41), we used the deictic "this" in "and does so like **this** - PSYCHO" to make the informational content of the musical gesture at-issue.

[13] In this paradigm, we embedded complex music in language and uncovered rich linguistic inferences. A non-trivial extension of our paradigm may in future test if such logical inferences can arise in purely musical environments.

(41) **Control premise**
If the dog comes to her and does so like this - PSYCHO, Marie will cross the street.

Original stimulus

Si le chien s' approche d' elle et qu' il le fait comme ça - PSYCHO, *Marie*
If the dog self approaches to her and that it-NOM. it-ACC. does like this - PSYCHO, Marie
changera de trottoir.
changes of sidewalk.

The use of "this" now makes the informational content of the musical gesture at-issue and thus targeted by the conditional "if", leading to the inference in (42) that the dog will not necessarily look vicious. Just like Tieu *et al.* (2019), we opted for a simpler formulation of the weaker control inference in (42) instead of the more convoluted exact negation of the target inference, "It is not the case that if the dog comes to Mary, it will look threatening."

(42) **Baseline inference**
If the dog comes to Mary, it won't necessarily look vicious.

3.3.3 Results

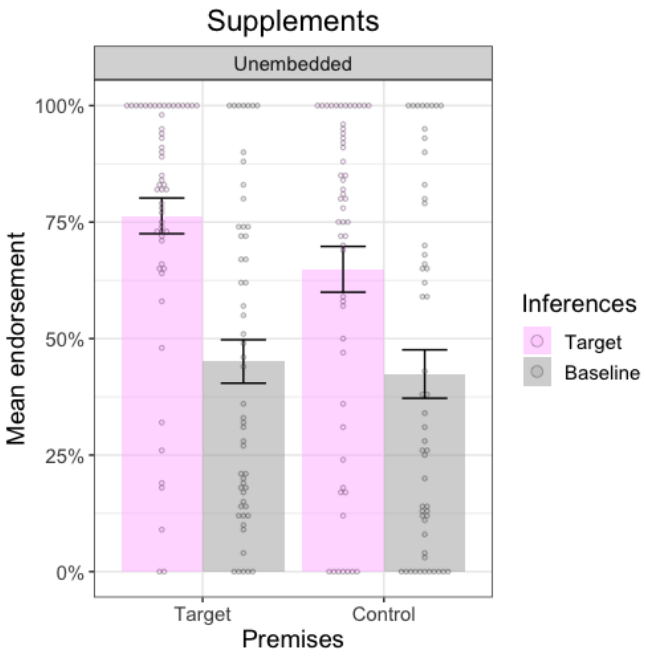

Figure 4: Mean endorsement rate for supplements

As explained in Appendix II, we are here interested in the interaction between inference type (supplement *vs* no supplement) and premise type (pro-speech gesture *vs* 'like this' control): to ensure that the stronger endorsement of the target inference containing the supplemental information we observe for the target premise was not due to a default preference for this kind of inference regardless of the premise, we contrasted this premise with a control in which the musical gesture was made at-issue through the use of 'like this'. Despite the higher endorsement of the target inference in response to the target premise, the interaction between the two factors was not significant (χ^2 = 1.5, p = 0.2). In Tieu *et al.* (2019), only a marginal interaction was found for supplements with visual animations. This lack of interaction might be due to the simplification of the negated inference used as a baseline, as we used "won't necessarily..." instead of the full negation "It is not the case that, if the dog comes to Mary, it will look vicious", which truth-conditions might well be different. We discuss a further difference with the original paradigm from Tieu *et al.* (2019) in the following subsection.

3.3.4 Discussion

The paradigm we used was slightly different from the test for supplements in Tieu *et al.* (2019), as we chose to introduce the content of the music explicitly in the context, while in Tieu *et al.* (2019), the gesture HIT mimicking someone boxing only occurred in the target sentences. We reasoned that introducing the content of the gesture would not affect its logical behavior. On the one hand, indeed, introducing the content of a pro-speech gesture would have jeopardized the validity of the results for inferences that can be easily lexically encoded, such as presuppositions under certain theories. On the other hand, because non-restrictive relative clauses are sentence-level constructions, there is not even the possibility that they are encoded in the meaning of a specific word.

Our paradigm did not aim at showing that subjects could derive a meaning from music but, rather, that they could treat a musical gesture as a non-restrictive relative clause when given its meaning. We thus introduced this meaning explicitly in the context, so that participants did not reject the target inference only because they perceived the excerpt as conveying a feeling other than fear. In Tieu et al's, the gesture HIT mimicking someone hitting someone else was used in lieu of a non-restrictive relative clause. It is quite unambiguous that HIT means something like *hit*. On the other hand, the meaning of the musical gesture, an excerpt from real complex film music, would have been hardly the same for all participants.[14] [15]

3.4 Homogeneity inferences

3.4.1 In language and from visual stimuli

Homogeneity inferences arise from definite plural noun phrases behaving universally in positive sentences, but existentially under negation (Križ, 2015; Križ & Spector, 2020; Spector, 2013; Löbner, 2000; Križ, 2016; Gajewski, 2005). Sentence (43)a has a universal reading: we understand that Mary found all of her presents. Under negation, however, the definite plural has an existential reading (i.e. "at least one"). Sentence (43)b is understood as *Mary did not find any of her presents*, and not as *Mary found some, but not all of her presents*, which involves the logical negation of "all of her present".

(43) a. Mary found her presents.
⇝ Mary found all of her presents.

b. Mary did not find her presents.
⇝ Mary did not find any of her presents.

Tieu *et al.* (2019) investigated whether pro-speech gestures and visual animations used in lieu of both a verb *and* a definite plural noun phrase could generate homogeneity inferences. For instance, in (44) and (45), the gesture TAKE-2-HANDED-RIGHT replaced both a verb meaning *take* and a definite plural. The definite component was realized by means of pointing, and the plural component was realized by means of the iteration of a gesture representing a coin, which is a common way of signaling plural in sign languages (Pfau & Steinbach, 2006). In the positive environment in (44), the pro-speech gesture triggers the universal inference that all coins were indeed taken.

(44) Sam will TAKE-2-HANDED-RIGHT.
⇝ Sam will take all of the coins.

Under negation, sentence (45) gives rise to an existential reading of the gesture TAKE-2-HANDED-RIGHT ("at least one coin") instead of the universal reading found in the positive environment ("all of the coins"). *Not*-TAKE-2-HANDED-RIGHT is interpreted as *not at least one coin*, i.e. no coin at all, rather than *not all of the coins*, i.e. some but not all of the coins.

[14] Although this music unambiguously conveys a feeling of fear, danger or suspense, there are many possible situations this music can refer to that would trigger such feelings, considering that music can indeed refer to several external non-musical situations sharing some structural and/or emotional properties (Schlenker, 2017, 2019b)

[15] As music is not generally used to directly communicate ideas or convey information about the world, investigating the intermediary case of onomatopoeias may bridge our findings with the results from Tieu *et al.* (2019). Since onomatopoeias are used in combination with language and are highly iconic, there are substantial reasons to believe that any inference type from the typology we replicated with musical gestures could just as well be replicated with onomatopoeias: since onomatopoeias occur naturally in lieu of words, we would expect them to display the same inferential behavior as musical gestures which do *not* occur naturally in lieu of words and are arguably more difficult to process. To verify that this was the case, we ran a similar experiment with onomatopoeias instead of musical stimuli. Results were not significant for all inference types (see Appendix I for details). However, the results were not significantly different across the two experiments, leaving open the possibility that the onomatopoeias experiment, which we ran on less participants, was lacking power.

(45) Sam will not TAKE-2-HANDED-RIGHT.
⇝ Sam will not take any coin.

While most research on homogeneity inferences focuses on the role of the noun phrase, more recent work has suggested that these inferences rather originate in the predicate associated with the noun phrase (Križ, 2019). We thus proceeded by case: in section 3.4.2.1, we report the results for stimuli in which the musical gesture replaced a noun phrase, while in section 3.4.3.1 we report a parallel paradigm but with the musical gesture in lieu of the predicate.

3.4.2 From music

3.4.2.1 Pro-speech music in lieu of the noun phrase

While the two components of the definite plural were visually realized in Tieu *et al.* (2019), it was not possible to realize them both in the auditory modality. The auditory counterpart of iteration, marking plural, was straightforward, and we used the iteration of a same musical note to mark plural. Crucially, however, we could not find an auditory counterpart of pointing to reproduce the definite component. We thus chose to focus on manipulating the iterative component only and on verifying that the musical gesture was interpreted as a definite plural.

In (46), HARP×3 stands for three repetitions of a same harp sound evoking three harp players in an orchestra, while FLUTES×3 stands for three repetitions of a same flute sound evoking three flute players in the same context.

(46) **NP Positive Context:** Every Thursday, the students of a music school are gathering, and the conductor chooses some of them to play for the evening concert. Tonight, three harp-players - HARP×3 and three flute-players FLUTE×3 are present.

We performed a first test in a positive environment, as shown in (47).

(47) **Target positive premise**
The conductor made HARP×3 play.

Original stimulus

Le chef d' orchestre a fait jouer HARP×*3.*
The chief of orchestra has made play HARP×3.

The musical gesture HARP×3 is expected to have a universal reading and be interpreted as *all harps*, leading to the inference in (48), which we expect to be significantly more endorsed than its negation, the baseline inference in (49).

(48) **Target inference**
All harps played.

(49) **Baseline inference**
Some but not all of the harps played.

Negation constitutes a crucial test to correctly explain the pattern observed in the positive environment, for two reasons. First, homogeneous expressions behave as existentials in negative environments. Second, an alternative explanation of the pattern we observe for the positive environment is that HARP×3 elicits a numeral, i.e. *exactly three harps*. Under negation, the two explanations come apart.

If HARP×3 is understood as a numeral, i.e. *three harps*, its negation *not*-HARP×3' should mean *not exactly three harps*, i.e. two or less, or four or more. This is compatible with the control in (52). If, by contrast, HARP×3 is understood as a definite plural, we expect it to be interpreted existentially under negation, i.e. *not*-HARP×3 is interpreted as *not at least one harp*, i.e. no harp at all.

(50) **Target negative premise**
The conductor did not make HARP×3 play.

Original stimulus

Le chef d' orchestre n' a pas fait jouer HARP×*3.*
The chief of orchestra NEG has not made play HARP×3.

If HARP×3 is interpreted as a definite plural, then we expect a higher endorsement of the homogeneous target sentence in (51) resulting from an existential reading of the musical gesture under negation, compared to the baseline inference in (52) resulting from a universal reading of the musical gesture (i.e., "The conductor did not make *all harps* play").

(51) **Target inference**
No harp played.

(52) **Baseline inference**
Some but not all of the harps played.

3.4.3 Results

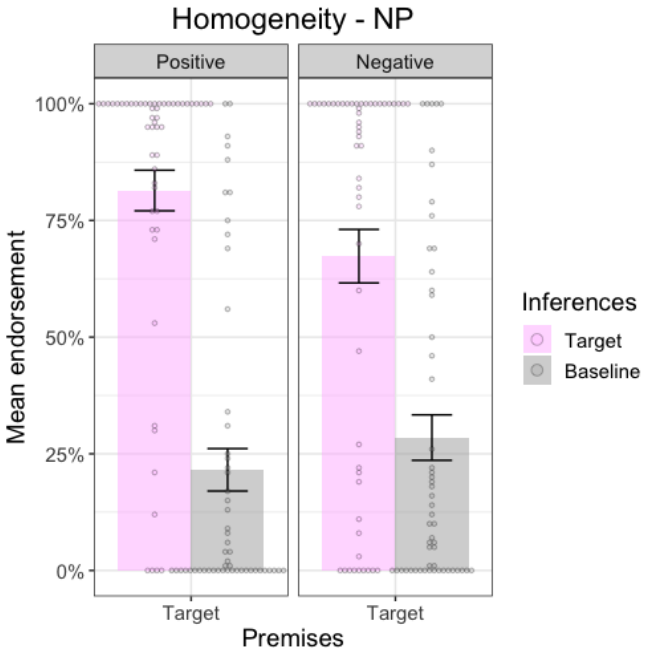

Figure 5: Mean endorsement rate for homogeneity (NP)

As explained in Appendix II, if homogeneity can be created by musical gestures standing for definite plurals, we expected an effect of inference type (homogeneous target *vs* non-homogeneous baseline): we expected the target homogeneous inference to be significantly more endorsed than its negation in both the positive and negative environments. This is what we find: the homogeneous inference ('all harps played') was preferred in response to the positive premise ($\chi^2 = 96$, $p < 0.001$), and similarly, the homogeneous inference ('no harp played') was preferred in response to the negative premise ($\chi^2 = 24$, $p < 0.001$). This suggests that it is possible to generate homogeneity inferences with musical gestures in lieu of definite plurals.[16]

3.4.3.1 Pro-speech music in lieu of the predicate

Recent work suggested that homogeneity inferences might not be systematically linked to a definite plural noun phrase, but rather to the kind of predicate associated with the noun phrase (Križ, 2019). If definite noun phrases were responsible for homogeneity inferences, it would be difficult to explain how non-homogeneous readings arise from sentences like (53).

[16] As pointed out by the Editor, instead of considering whether the premise triggered a homogeneous or a non-homogeneous reading, we could have shown that the interpretation of the musical gesture flips from universal in the positive environment to existential in the negative environment. In this case, we would have been interested in the interaction between the inference type factor, and the environment (positive *vs* negative). We display this alternative analysis in Appendix II, and show that it leads to a significant interaction between the two factors, indicating that the reading did change from universal to existential once embedded under negation.

(53) The students are numerous.
a. ?? Each of the students is numerous.

Križ (2019) argues that predicates, rather than bare plurals, are responsible for the presence or absence of homogeneity . While in Tieu *et al.* (2019), the test for homogeneity involved the gesture TAKE_ALL encompassing both the predicate TAKE and a definite plural noun phrase THE_COINS, the test we presented in section 3.4.2.1 only involved a musical gesture replacing the noun phrase.

Since there is the theoretical possibility that homogeneity comes from the predicate, we also tested whether sentences in which pro-speech music is inserted in lieu of the *predicate* can give rise to homogeneity inferences as well. In (54), the musical gesture MARSEILLAISE featuring a children's choir singing the French national anthem was used to evoke the action of singing.

(54) **Positive Context (predicate)**
The students of a class are learning to sing the Marseillaise - MARSEILLAISE.

We applied the same protocol as in section 3.4.2.1. We performed a first test in a positive environment, where we expected sentence (55) to give rise to the universal inference in (56).

(55) **Target positive premise**
On the 14th of July, the class has MARSEILLAISE.

Original stimulus

Le jour du 14 juillet, la classe a MARSEILLAISE.
The day of 14th July, the class has MARSEILLAISE.

(56) **Target inference**
The children in the class all sang the Marseillaise.

Just as in 3.4.2.1, the target inference in (56) needs to be contrasted with the non-universal baseline inference in (57).

(57) **Baseline inference**
Some but not all of the children in the class sung the Marseillaise.

We then performed a second test in a negative environment, i.e. the predicate MARSEILLAISE was in the scope of negation, as in (58).

(58) **Target negative premise**
On the 14th of July, the class did not MARSEILLAISE.

Original stimulus

Le jour du 14 juillet, la classe n' a pas MARSEILLAISE.
The day of 14th July, the class NEG has not MARSEILLAISE.

The negation of the musical predicate (*not*-MARSEILLAISE) now applies homogeneously to each member of the set denoted by the group singular noun phrase "the class", leading to the inference in (59) that no child in the class sang the Marseillaise, which we expected to be highly endorsed.

(59) **Target inference**
None of the children in the class sang the Marseillaise.

By contrast, we expected a low endorsement of the logical negation of this inference, i.e. the weaker baseline inference in (60) that some but not all of the children in the class sang the Marseillaise.

(60) **Baseline inference**
Some but not all of the children in the class sung the Marseillaise.

In this paradigm, none of the tested inferences involve a definite plural noun phrase, so that the effect of the musical predicate could be isolated if there is one. If the sentence gives rise to a homogeneity inference, we expect a strong endorsement of the target inferences (i.e. a universal reading of the sentence in the positive environment, and an existential one under negation) compared to the baseline inference.

An intuitive objection is that these inferences are not due to the predicate but to the group singular noun phrase "the class". In this case, however, the homogeneity inference originates in the noun phrase and thus our paradigm presented in 3.4.2.1 provides the relevant evidence. If, on the other hand, homogeneity originates in the predicate, the results presented in this section provide evidence that pro-speech music predicates can give rise to homogeneity inferences.

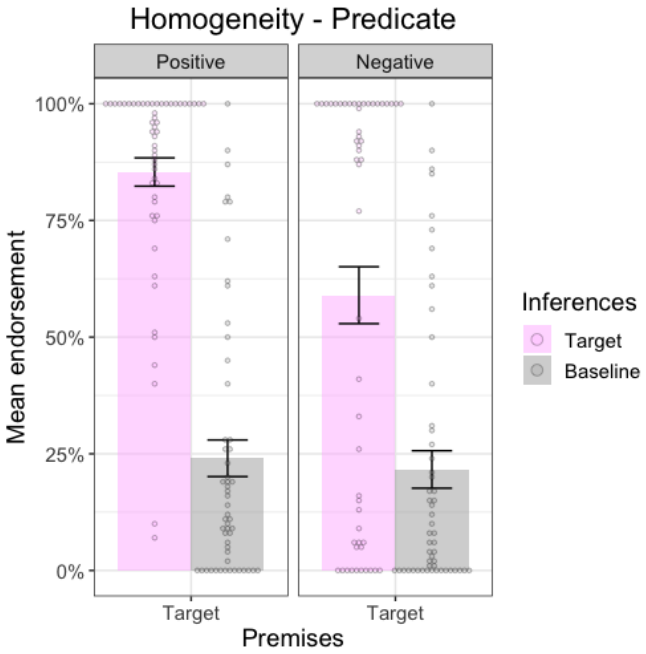

Figure 6: Mean endorsement rate for homogeneity (predicate)

The results show a significant effect of inference type in both the positive ($\chi^2 = 96$, $p < 0.001$) and the negative environments ($\chi^2 = 24$, $p < 0.001$). This suggests that it is possible to give rise to homogeneity inferences by means of a musical gesture replacing the predicate instead of the definite plural noun phrase.

4 Iconicity controls

4.1 Experimental paradigm

As mentioned in section 1, an alternative explanation of our data is that musical gestures gave rise to the observed inferences simply because they are systematically translated into words. Note that Tieu *et al.* faced the very same concern with pro-speech gestures and visual animations, which was also addressed by including iconicity controls throughout the experiment.

To rule out this alternative explanation, we included 8 iconicity trials randomly displayed throughout the experiment. We manipulated these stimuli to check whether participants were able to interpret their modification iconically. To this end, we used two different realizations of an upward scale, a normal and a fast one, to refer to a hiker climbing either slowly or quickly. We used two other realizations, a very long musical scale and a shorter one, to evoke the height of the mountain, in the same context as the one used in 3.2. The target sentences were then paired up with a matching or a mismatching inference. If participants were able to retrieve the expected iconic

informational content from the musical gesture, we expected a significantly higher endorsement of the matching inference for the target premise with respect to the mismatching inference.

For the first iconicity control in (61), LONG_UPWARD_SCALE stands for a C major scale played over a wide pitch range of four octaves (i.e., four full upward scales, from middle pitch to very high), used to evoke a very high mountain.

(61) **Target premise**
The next hiker will LONG_UPWARD_SCALE.

Original stimulus

Le	*prochain*	*randonneur*	*va*	LONG_UPWARD_SCALE.
The	next	hiker	will	LONG_UPWARD_SCALE.

We expected participants to be able to (i) perceive the difference between the two realizations of the upward scale and (ii) interpret them iconically in terms of the height of the mountain in the context. Consequently, we expected a high endorsement of the inference containing the matching iconic component in (62).

(62) **Matching inference**
The mountain the hiker is about to climb is high.

In (63), UPWARD_SCALE stands for the same scale played at the same speed but over a narrower pitch range of two octaves (i.e., two full upward scales only, from middle pitch to moderately high), to evoke a mountain of average height. Note that this musical gesture was the one used in 3.2 to test for presupposition projection. Our prediction was that the longer the scale, the higher the inferred height of the mountain.

(63) **Control premise**
The next hiker will UPWARD_SCALE.

Original stimulus

Le	*prochain*	*randonneur*	*va*	UPWARD_SCALE.
The	next	hiker	will	UPWARD_SCALE.

We expected that this neutral realization of the scale would convey information about the mountain being of average height, and thus a high endorsement of the inference in (64), in which the iconic component allowing subjects to assess the mountain to be very high was absent.

(64) **Mismatching inference**
The mountain the hiker is about to climb is not very high.

For the second iconicity control, we contrasted the baseline realization of the scale UPWARD_SCALE in (67) evoking a neutrally fast hike with SLOW_UPWARD_SCALE in (65), in which the pitch range remained the same but the speed was slowed down four times to evoke a slower climbing of the mountain.

(65) **Target premise**
The next hiker will SLOW_UPWARD_SCALE.

Original stimulus

Le	*prochain*	*randonneur*	*va*	SLOW_UPWARD_SCALE.
The	next	hiker	will	SLOW_UPWARD_SCALE.

Similarly, participants should be able to both perceive and interpret the difference between the two realizations iconically in terms of the hiking speed, and we expected a high endorsement of the inference in (66) containing the iconic component.

(66) **Matching inference**
The hiker will climb the mountain slowly

(67) **Control premise**
The next hiker will UPWARD_SCALE.

Original stimulus

Le	*prochain*	*randonneur*	*va*	UPWARD_SCALE.
The	next	hiker	will	UPWARD_SCALE.

The iconic component allowing for the inference about a high hiking speed was absent from this control premise, which we expected to give rise to the inference in (68).

(68) **Mismatching inference**
The hiker will climb the mountain fast.

We found a significant effect of inference type ($\chi^2 = 63$, $p < 0.001$ for the first iconicity control, $\chi^2 = 28$, $p < 0.001$ for the second). This means that listeners were able to both perceive subtle musical changes and interpret them iconically.

4.2 Discussion

We have shown that musical gestures convey fine-grained iconic information. Specifically, the pitch range of the stimulus was interpreted as the height of the mountain, and the tempo at which the upward scale was played was interpreted as the hikers' speed. It is very unlikely that these musical gestures were merely translated in words, because all these iconic dimensions are absent from reasonably simple translations. For instance, the information about height conveyed by UPWARD_SCALE is absent from a simple translation as "go up". It could still be, however, that more complex translations are involved. These would have to be extraordinarily rich: "The hiker will LONG_UPWARD_SCALE" would have to be translated as something like "The hiker will climb a mountain of such-and-such height at such-and-such a speed...". However, even if pro-speech music were indeed translated into such long chains of words, the music-as-translation theory makes wrong predictions on the projection patterns of these iconic dimensions. Consider a translation of the target premise which, for ease, we simplify as (69).

(69) The hiker will climb up the mountain slowly.

The music-as-translation theory predicts that all pieces of fine-grained information iconically conveyed be *systematically at-issue* when not *lexically* encoded as non-at-issue. Any musical gesture conveying fine-grained iconic information should therefore systematically give rise to a scalar implicature. That is, SLOW_UPWARD_SCALE would mean *climb slowly*, thus its negation *not*-SLOW_UPWARD_SCALE should give rise to the scalar implicature that there is still some climbing involved, given that the more informative *not*-UPWARD_SCALE was not uttered. This prediction seems right if we look at the negation of (69), as in (70).

(70) The hiker will not climb up the mountain slowly.
⇝ The hiker will not climb up the mountain slowly, but will still climb up the mountain.

The premise in (70) does not, however, trigger the same inference when using the musical gesture LONG_UPWARD_SCALE.

(71) The hiker will not LONG_UPWARD_SCALE.
a. ⇝ The hiker will not climb up the mountain slowly, but will still climb up the mountain.
b. ⇝ The hiker will not climb the mountain at all, but if she did, she would have done so slowly.

A first possible reading of the sentence in (71) is the same scalar implicature found in (71)a. However, another possible reading is (71)b, where the iconic information paraphrasable as "slowly" is not at-issue anymore. While this reading is not the most obvious, it becomes completely salient in context. Consider the following example:

(72) **Context:** Mary and John are playing a snail race, and they want their snails to climb all sorts of obstacles.
Mary tells John: "Your snail will not SLOW_UPWARD_SCALE."
⇝ John's snail will not climb the next obstacle but if it had climbed it, it would have done so slowly.
↛ John's snail will not climb the next obstacle slowly, but will still climb it.

This inference is a typical cosupposition (Schlenker, 2018a) of the form *x did not Q, but if x did Q, x would have Qed like this*. Crucially, it is not triggered in the purely linguistic example in (70), as shown in (73).

(73) The hiker will not climb up the mountain slowly.
↝ The hiker will not climb the mountain at all, but if she did, she would have done so slowly.

If the musical gesture is only translated into words and no other non-translatable iconic enrichment is involved, then there is at least one available reading which cannot be explained. Consequently, there is no real alternative to the conclusion that there is a non-verbal component to pro-speech music. It is worth noting that similar inferences were found to be triggered by co-speech sounds and auditory animations (Tieu *et al.*, 2018), which make non-at-issue contributions to the sentence.

5 Conclusion

The purpose of this study was to replicate the inferential typology tested on pro-speech gestures and visual animations of inferences in Tieu *et al.* (2019) with pro-speech music. Participants behaved systematically in response to novel stimuli, suggesting that a general cognitive algorithm is responsible for the appropriate arrangement of content, rather than a purely linguistic procedure relying on word meaning. Specifically, our results show that pro-speech music can give rise to four types of inferences.

We showed that musical gestures can assume different levels of informativity and compete, resulting in the computation of scalar implicatures (cf: section 3.1).

Moreover, we showed that pro-speech music can convey presuppositions. Presupposed content projected under question (cf. section 3.2), but did not project universally under “none”. While this difference with Tieu *et al.* (2019) weakens our hypothesis a bit, it does not constitute negative evidence given the complex behavior of presuppositions under “none". Testing simple negation will provide a helpful test to show whether the observed behavior under “none" is specific to “none" or due to general complex projection patterns of pro-speech music.

Conditional inferences typical of supplements, i.e. non-restrictive relative clauses, can also arise from rich pro-speech music (cf. section 3.3). However, although the target inference with the supplemental information was more highly endorsed than the control inference without this information, we did not find the expected interaction between premise type and inference type. As Tieu *et al.* (2019) reported a marginal interaction in the case of pro-speech visual animations, further research will be needed to understand whether abstractness might weaken the triggering of supplements.

In section 3.4, we showed that pro-speech music can give rise to homogeneity inferences, both when replacing a definite plural noun phrase and when inserted in lieu of a predicate. Section 4 finally illustrated why it is very unlikely that these results can be explained by the mere translation of musical gestures into words.

Together, these findings suggest that the mechanisms responsible for these inferences and for the division of meaning along different slots of the inferential typology are (i) productive and (ii) general purpose. Indeed, (i) participants were able to draw the expected inferences in absence of previous exposure with the musical gestures, and (ii) they were able to operate logical computations on auditory, non-linguistic stimuli.

Finally, the question of the existence of a music semantics, recently raised in Schlenker (2017, 2019b) (a.o.), is a very exciting and promising one, but our results do not allow us to make any claim about musical meaning. However, we think that the evidence we provide about the interaction between language structure and musical content constitutes a first step to better understand what musical meaning is and how it behaves logically. In particular, the extent to which some of the observed inferences may arise in music alone, i.e. not embedded in language, is an inspiring question for future research.

References

Abbott, B. (2000) Presuppositions as nonassertions. *Journal of Pragmatics* **32**, 1419–1437.

Abrusán, M. (2010) Triggering verbal presuppositions **20**, 684–701.

Abrusán, M. (2011) Predicting the presuppositions of soft triggers. *Linguistics and Philosophy* **34**, 491–535.

Abusch, D. (2002) Lexical Alternatives as a Source of Pragmatic Presuppositions. *Semantics and Linguistic Theory* **12**, 1.

Abusch, D. (2010) Presupposition Triggering from Alternatives. *Journal of Semantics* **27**, 37–80.

Barr, D.J., Levy, R., Scheepers, C. & Tily, H.J. (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**, 255–278.

Beaver, D. (1994) When variables don't vary enough. *Semantics and linguistic theory*, vol. 4, pp. 35–60.

Beaver, D.I. (2001) Presupposition and Assertion in Dynamic Semantics. *Stanford: CSLI Publications.* .

Breheny, R. (2019) Scalar Implicatures. *The Oxford Handbook of Experimental Semantics and Pragmatics* (eds. C. Cummins & N. Katsos), pp. 38–61, Oxford University Press.

Chemla, E. (2009) Presuppositions of quantified sentences: experimental data. *Natural Language Semantics* **17**, 299–340.

Chemla, E. & Schlenker, P. (2012) Incremental vs. symmetric accounts of presupposition projection: an experimental approach. *Natural Language Semantics* **20**, 177–226.

Chierchia, G. (2006) Broaden Your Views: Implicatures of Domain Widening and the "Logicality" of Language. *Linguistic Inquiry* **37**, 535–590.

Chierchia, G., Fox, D. & Spector, B. (2012) Scalar implicature as a grammatical phenomenon. *Semantics: An International Handbook of Natural Language Meaning*, De Gruyter Mouton, Berlin, claudia maienborn, klaus von heusinger, and paul portner edn.

Chierchia, G. & McConnell-Ginet, S. (2000) *Meaning and grammar: An introduction to semantics*. MIT press.

Chomsky, N. (1980) On Cognitive Structures and Their Development: A Reply to Piaget. *Language and Learning*, Routledge and Kegan Paul, London, piatelli-palmarini, m., ed. edn.

Chomsky, N. (1988) *Language and problems of knowledge: the Managua lectures*. No. 16 in Current studies in linguistics series, MIT Press, Cambridge, Mass.

Clark, H.H. (2016) Depicting as a method of communication. *Psychological Review* **123**, 324–347.

Ebert, C. & Ebert, C. (2014) Gestures, Demonstratives, and the Attributive/Referential Distinction. Handout of a talk given at Semantics and Philosophy in Europe (SPE 7).

Fox, D. (2007) Free Choice and the Theory of Scalar Implicatures. *Presupposition and Implicature in Compositional Semantics* (eds. U. Sauerland & P. Stateva), Palgrave Studies in Pragmatics, Language and Cognition, pp. 71–120, Palgrave Macmillan UK, London.

Fox, D. (2013) Presupposition projection from quantificational sentences: trivalence, local accommodation, and presupposition strengthening. *MIT web domain* .

Fox, D. & Katzir, R. (2011) On the characterization of alternatives. *Natural Language Semantics* **19**, 87–107.

Gajewski, J. (2005) *Neg-raising: Polarity and Presupposition. MIT dissertation.* Ph.D. thesis, MIT.

George, B.R. (2008) A New Predictive Theory of Presupposition Projection. *Semantics and Linguistic Theory* **18**, 358–375.

Geurts, B. (1998) The Mechanisms of Denial. *Language* **74**, 274–307.

Geurts, B. (1999) *Presuppositions and pronouns*. No. v. 3 in Current research in the semantics/pragmatics interface, Elsevier, Amsterdam ; New York, 1st ed edn.

Geurts, B. (2010) *Quantity Implicatures*. Cambridge University Press, Cambridge.

Geurts, B. & van Tiel, B. (2016) When "All the Five Circles" are Four: New Exercises in Domain Restriction. *Topoi* **35**, 109–122.

Goldin-Meadow, S. & Brentari, D. (2017) Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences* **40**, e46.

Grice, H. (1975) Logic and conversation. *Syntax and Semantics 3: Speech Acts*, pp. 41 – 58, peter cole and jerry morgan edn.

Heim, I. (1988) *The semantics of definite and indefinite noun phrases*. Outstanding dissertations in linguistics, Garland Pub, New York.

Heim, I. (1990) Presupposition Projection. *Presupposition, Lexical Meaning and Discourse Processes: Workshop Reader*, p. 33, R. van der Sandt, University of Nijmegen.

Heim, I. (1992) Presupposition Projection and the Semantics of Attitude Verbs. *Journal of Semantics* **9**, 183–221.

Heim, I. (2002) On the Projection Problem for Presuppositions. *Formal Semantics*, pp. 249–260, John Wiley & Sons, Ltd.

Heim, I. & Kratzer, A. (1998) *Semantics in generative grammar*. No. 13 in Blackwell textbooks in linguistics, Blackwell, Malden, MA.

Horn, L. (1972) *On the semantic properties of the logical operators in English. Ph.D. thesis*. Ph.D. thesis, University of California at Los Angeles.

Kadmon, N. (2001) *Formal pragmatics: semantics, pragmatics, presupposition, and focus*. Blackwell, Malden, Mass., 1. publ edn.

Katzir, R. (2007) Structurally-defined alternatives. *Linguistics and Philosophy* **30**, 669–690.

Križ, M. (2015) *Aspects of homogeneity in the semantics of natural language.* PhD Thesis, University of Vienna.

Križ, M. (2016) Homogeneity, Non-Maximality, and *all*. *Journal of Semantics* **33**, 493–539.

Križ, M. (2019) Homogeneity effects in natural language semantics. *Language and Linguistics Compass* **13**.

Križ, M. & Spector, B. (2020) Interpreting plural predication: homogeneity and non-maximality. *Linguistics and Philosophy* .

Levinson, S.C. (2000) *Presumptive meanings: the theory of generalized conversational implicature*. Language, speech, and communication, MIT Press, Cambridge, Mass.

Löbner, S. (2000) Polarity in natural language: predication, quantification and negation in particular and characterizing sentences. *Linguistics and Philosophy* **23**, 213–308.

Mandelkern, M. (2016) A note on the architecture of presupposition. *Semantics and Pragmatics* **9**, 13–1–24.

Mayr, C. & Sauerland, U. (2016) Accommodation and the strongest meaning hypothesis. *Preproceedings of the Amsterdam Colloquium* .

McNeill, D. (2005) *Gesture and thought*. University of Chicago Press, Chicago.

Pfau, R. & Steinbach, M. (2006) Pluralization in sign and in speech: A cross-modal typological study. *Linguistic Typology* **10**.

Potts, C. (2004) *The Logic of Conventional Implicatures*. Oxford University Press.

Potts, C. (2015) Presupposition and Implicature. *The Handbook of Contemporary Semantic Theory*, pp. 168–202, John Wiley & Sons, Ltd.

Sauerland, U. (2004) Scalar Implicatures in Complex Sentences. *Linguistics and Philosophy* **27**, 367–391.

Sauerland, U. (2012) The Computation of Scalar Implicatures: Pragmatic, Lexical or Grammatical?: Computation of Scalar Implicatures. *Language and Linguistics Compass* **6**, 36–49.

Schlenker, P. (2008) Presupposition Projection: the New Debate. *Semantics and Linguistic Theory* **18**, 655–693.

Schlenker, P. (2010) Presuppositions and Local Contexts. *Mind* **119**, 377–391.

Schlenker, P. (2016) The semantics–pragmatics interface. *The Cambridge Handbook of Formal Semantics* (eds. M. Aloni & P. Dekker), pp. 664–727, Cambridge University Press, Cambridge.

Schlenker, P. (2017) Outline of Music Semantics. *Music Perception: An Interdisciplinary Journal* **35**, 3–37.

Schlenker, P. (2018a) Gesture projection and cosuppositions. *Linguistics and Philosophy* **41**, 295–365.

Schlenker, P. (2018b) Iconic pragmatics. *Natural Language & Linguistic Theory* **36**, 877–936.

Schlenker, P. (2019a) Gestural semantics: Replicating the typology of linguistic inferences with pro- and post-speech gestures. *Natural Language & Linguistic Theory* **37**, 735–784.

Schlenker, P. (2019b) Prolegomena to Music Semantics. *Review of Philosophy and Psychology* **10**, 35–111.

Schlenker, P. (2020) Gestural grammar. *Natural Language & Linguistic Theory* **38**, 887–936.

Schlenker, P. (2021) Triggering Presuppositions. *Glossa: a journal of general linguistics* **6**.

Simons, M., Tonhauser, J., Beaver, D. & Roberts, C. (2010) What projects and why. *Semantics and Linguistic Theory* **20**, 309.

Spector, B. (2007a) Aspects of the Pragmatics of Plural Morphology: On Higher-Order Implicatures. *Presupposition and Implicature in Compositional Semantics* (eds. U. Sauerland & P. Stateva), pp. 243–281, Palgrave Macmillan UK, London.

Spector, B. (2007b) Scalar Implicatures: Exhaustivity and Gricean Reasoning. *Questions in Dynamic Semantics* (eds. M. Aloni, A. Butler & P. Dekker), pp. 225–249, BRILL.

Spector, B. (2013) Homogeneity and plurals: From the strongest meaning hypothesis to supervaluations. .

Stalnaker, R. (1974) Pragmatic Presuppositions. *Context and Content* (ed. R. Stalnaker), pp. 47–62, Oxford University Press.

Stalnaker, R. (1975) Presuppositions. *Contemporary Research in Philosophical Logic and Linguistic Semantics* (eds. D. Hockney, W. Harper & B. Freed), pp. 31–41, Springer Netherlands, Dordrecht.

Sudo, Y., Romoli, J., Hackl, M. & Fox, D. (2012) Presupposition Projection Out of Quantified Sentences: Strengthening, Local Accommodation and Inter-speaker Variation. *Logic, Language and Meaning* (eds. M. Aloni, V. Kimmelman, F. Roelofsen, G.W. Sassoon, K. Schulz & M. Westera), Lecture Notes in Computer Science, pp. 210–219, Springer, Berlin, Heidelberg.

Team, R.C. (2016) R: A Language and Environment for Statistical Computing (R Foundation for Statistical Computing, Vienna), Version 3.4.3.

Tieu, L., Pasternak, R., Schlenker, P. & Chemla, E. (2018) Co-speech gesture projection: Evidence from inferential judgments. *Glossa: a journal of general linguistics* **3**, 109.

Tieu, L., Schlenker, P. & Chemla, E. (2019) Linguistic inferences without words. *Proceedings of the National Academy of Sciences* **116**, 9796–9801.

Tonhauser, J., Beaver, D., Roberts, C. & Simons, M. (2013) Toward a Taxonomy of Projective Content. *Language* **89(1):66-109**.

Van Der Sandt, R.A. (1992) Presupposition Projection as Anaphora Resolution. *Journal of Semantics* **9**, 333–377.

van Rooij, R. & Schulz, K. (2004) Exhaustive Interpretation of Complex Sentences. *Journal of Logic, Language and Information* **13**, 491–519.

Zehr, J., Bill, C., Lyn, T., Jacopo, R. & Florian, S. (2016) Presupposition projection from the scope of None: Universal, existential, or both? *Semantics and Linguistic Theory* **26**, 754–774.
unsrt

Appendix I: Musical gestures and vocal gestures

As mentioned in section 3.3, we ran a parallel experiment on a different pool of participants using pro-speech onomatopoeias, i.e. iconic vocal sounds replacing one or several words in sentences, that we call 'vocal gestures', instead of musical gestures. The same four inference types were tested: scalar implicatures, presuppositions, supplements and homogeneity inferences, using paradigms and stimuli that were either perfectly analogous to the ones described throughout this paper, or very slightly different when vocal counterparts to musical stimuli could not be found. For instance, wherever an upward scale was used in our stimuli to evoke a rise in space to test for presuppositions, we used a whistle rising in frequency, and wherever a choir singing the French national anthem was used to evoke the action of singing to test for homogeneity inferences, we used a similar stimulus where the very same song was whistled instead of sung. Wherever a drum was used to evoke someone boxing to test for scalar implicatures, we used the onomatopoeia BOOM vocally pronounced.

We found a difference in the results collected from the experiment on musical gestures and the one on vocal gestures, which is surprising as most paradigms were perfectly symmetric. However, this experiment was ran on a smaller pool than the musical gestures experiment, and the lack of systematically significant effects of the inference type might thus be explained by a lack of power. We found indeed that the differences in the endorsement rates across both experiments were mainly not or marginally significant themselves, i.e. the responses given for the experiment on vocal gestures were not significantly different from the ones given for the experiment on musical gestures. The table below summarizes the comparisons of the distribution of the data between both experiments. For each inference type, we computed the interaction between inference type and the *Experiment* factor, whose two levels corresponded to the two experiments. [17]

Inference type	**Environment**	***p*-value for the [*Inference type x Experiment*] interaction**	**Significance interpretation**
Scalar implicatures	Positive	0.74	No interaction
	Negative	0.08	Marginally significant interaction
Presuppositions	Question	0.61	No interaction
	None	0.43	No interaction
Supplements		< 0.001	Significant interaction (as expected)
Homogeneity	NP	0.05	Low interaction
	Predicate	0.29	No interaction

Comparison between musical gestures and vocal gestures experiments

The table above displays the figures assessing the significance of the difference in distribution of the data collected across experiments. None of the data subsets were significantly different across both experiments, except for supplements, for which we found no contrast in endorsement due to an unexpected interpretation of the vocal stimulus.

[17] The detailed analyses and statistical scripts to compute interactions are available in the RESULTS folder at https://osf.io/hw45u/?view_only=89f983db777f49e9a6f5b41b3dea60d6

Appendix II: Statistical models

In this appendix, we provide the generalized linear mixed-effects models we used to analyze the data. They were all inspired by that of (Tieu *et al.*, 2019), whose analysis script is open source. For each model, we justify the structure of the model and in particular its random effects structure by the design (as recommended in Barr *et al.* (2013)). We also provide the coefficients found for each model. For the sake of clarity, we report the statistical models in a raw fashion (i.e. the R code lines). Useful information about the syntax of R, which can be helpful to read these formulas, can be found online and are accessible at this link:
https://github.com/clayford/LMEMInR/blob/master/lme4_cheat_sheet.Rmd.

Scalar implicatures

In the scalar implicatures paradigms, two main factors could have had an effect on the endorsement of the target inferences: the gesture factor (GestureC), which two levels correspond to the target and control premises contrasting two alternative realizations of the drum sound mimicking someone boxing, DRUM×1 and DRUM×2; and the inference factor (InferenceC), which two levels correspond to the target and baseline inferences contrasting two possible interpretations of the the drum sounds ('some' vs 'a lot' in the positive environment; 'some' vs 'none' in the negative environment).

Here, we were interested in the `GestureC * InferenceC` interaction to ensure that the contrast found for the target premise was not due to a default bias in endorsement for one inference over the other, by comparing the difference between the endorsement of the target and that of its negation for the target premise and the same difference for the control premise. We did *not* include the environment factor in the model (positive *vs* negative) because there was no theoretical reason to expect a difference in these interactions between the positive and the negative environment. `GestureC`, `InferenceC`, and their interaction were thus used as fixed effects in our model, while this interaction *by subject* was used as a random effect, accounting for the variability across participants in (i) the interpretation of the premise, (ii) the interpretation of the inference, and (iii) the interaction between (i) and (ii).

Model 1

```
value ~ GestureC * InferenceC + (1 + GestureC * InferenceC|SubjID)
```

This first model failed to converge. We followed the same procedure as (Tieu *et al.*, 2019) in simplifying the random effects structure, and removed the interaction between the two factors `GestureC` and `InferenceC` from the random effects as follows:

Simplification of random effects structure (as in (Tieu *et al.*, 2019))

```
value ~ GestureC * InferenceC + (1 + GestureC + InferenceC|SubjID)
```

Environment	(Intercept)	GestureC	InferenceC	GestureC:InferenceC
Positive	52.60377	-0.79245	-0.07547	116.60377
Negative	57.005	1.557	7.575	44.698

Presuppositions

There was no theoretical reason for predicting an interaction in presuppositional behavior between question formation and projection under 'none', as the two projection tests are independent. Inferences from questions and from 'none' were thus analyzed separately. We are rather interested in the effect of the inference type (presupposition vs no presupposition) on the responses.

To ensure that the inference containing presupposition was not due to a by-default preference for this kind of inference, the endorsement of the presupposition was contrasted with that of its negation (baseline inference), as a control. The model thus used the inference factor InferenceC, which levels correspond to the presupposition/no

presupposition, as a fixed effect, and this same factor *by subject* was used as a random effect, to capture the possibility that each participants may simply tend to endorse presuppositions differently.

Model 1
`value ~ InferenceC + (1 + InferenceC|SubjID)`

This first model failed to converge, leading us to simplify the random effect structure by only keeping the SubjID factor as a random effect, to capture the variability in intercept across participants:

Simplification of random effects structure
`value ~ InferenceC + (1 | SubjID )`

Environment	(Intercept)	InferenceC
Question	49.48	-15.60
'none'	51.65	2.74

Supplements

The case for supplements is symmetric to the paradigm for scalar implicatures in 5, where both the gesture factor GestureC, describing the two types of premise (the target premise where the musical gesture is expected to be interpreted as a non-restrictive relative clause, and the control premise where the musical gesture is made at-issue by using 'like this') and the inference factor InferenceC, describing the two types of inference (supplemental or not) were used as fixed effects in the model, while the interaction *by participant* was used as a random effect in the maximal model below:

Here, we are interested in the interaction between the Gesture factor, which two levels represent the two forms of sentences (with a pro-speech gesture standing for a non-restrictive relative clause, and with the deictic 'like this') and the Inference factor, which two levels correspond to the supplemental inference (if X, then X would have happened in a certain way) and the inference without the supplemental information (if X, then X would not have necessarily happened in this same way). As the control inference using 'not necessarily' was not the exact negation of the target inference for reasons of simplicity, we had no strong prediction as to how differently the target and control inference would be endorsed for the control premise using 'like this'; but we expected this difference to be important for the target inference using a post-speech musical gesture (without 'like this') which was expected to trigger a supplemental inference just as post-speech gestures do (Tieu *et al.*, 2019; Schlenker, 2018a). We thus expected an interaction between the two factors, with a higher difference in endorsement between the target inference and the baseline inference in response to the target premise than in response to the control premise.

Model 1
`value ~ GestureC * InferenceC + (1 + GestureC * InferenceC|SubjID)`

This first model failed to converge, leading us to simplify as before the random effect structure by removing the interaction:

Simplification of random effects structure 1
`value ~ GestureC * InferenceC + (1 + GestureC + InferenceC|SubjID)`

This second model converged but did not allow for interaction testing (removing the interaction from the model prevented the model from converging), leading us to simplify the random effect structure even more:

Simplification of random effects structure 2
`value ~ GestureC*InferenceC+(1+InferenceC|SubjID)`

(Intercept)	GestureC	InferenceC	GestureC:InferenceC
57.175	-7.085	-26.858	8.774

Homogeneity inferences - NP and Predicate

The case for homogeneity is analogous to that of presuppositions in 5, where only inference type was used as a fixed effect in the model, while the effect of inference type **by participant** (testing whether each participant had a personal tendency to endorse the inferences) was used as a random effect. As we were interested in whether the homogeneous ('all' or 'none') inference would be preferred to the non-homogeneous one in each environment (positive and negative), we are here only interested in the effect of inference type (homogeneous *vs* non-homogeneous).

Model 1

```
value ~ InferenceTypeC + (1 + InferenceTypeC|SubjID)
```

This model did not converge, so we decided, as with presuppositions, to go with the most minimal random effect structure only accounting for the difference in intercepts between participants:

Simplification of random effects structure

```
value ~ InferenceTypeC + (1|SubjID)
```

Hypothesis tested	Environment	(Intercept)	InferenceC
NP	Positive	51.49	-59.85
NP	Negative	47.91	-38.87
Predicate	Positive	54.71	-61.34
Predicate	Negative	40.31	-37.34

Note that in our analysis, we were interested in inference type, i.e. whether the homogeneous inference was significantly more endorsed than its non-homogeneous counterpart. However, as the editor pointed out, it would also have made sense to analyze the inference type in terms of *universal* or *existential*. As described in the paradigm in 3.4, we expected musical gestures (that were punctuated repetitions) to have a universal reading in the positive environment ('all harps') and to get an existential reading in the negative environment ('at least one harp', i.e. 'some harps'), just as the reading of gestures and visual animations were shown to flip from universal to existential depending on the context in (Tieu *et al.*, 2019). If we perform such an analysis, we are interested in the interaction between environment (positive *vs* negative) and inference type (understood as universal vs existential, and not as homogeneous vs non-homogeneous anymore). We plot the results with this new analysis and report the significance and coefficients of what the model would have been in this situation below:

Model 1

```
value ~ InferenceTypeC * Environment + (1 + InferenceTypeC * Environment|SubjID)
```

This model failed to converge, leading to a simplification of the random effects structure as follows:

Simplified model

```
value ~ InferenceTypeC * Environment + (1 + InferenceTypeC + Environment|SubjID)
```

The comparison of this model to the same model without the interaction between inference type and environment showed a significant difference between the two models ($\chi^2 = 84$, $p < 0.001$). This supports the idea that the reading of a same musical gesture flips under negation, which is consistent with our first analysis.

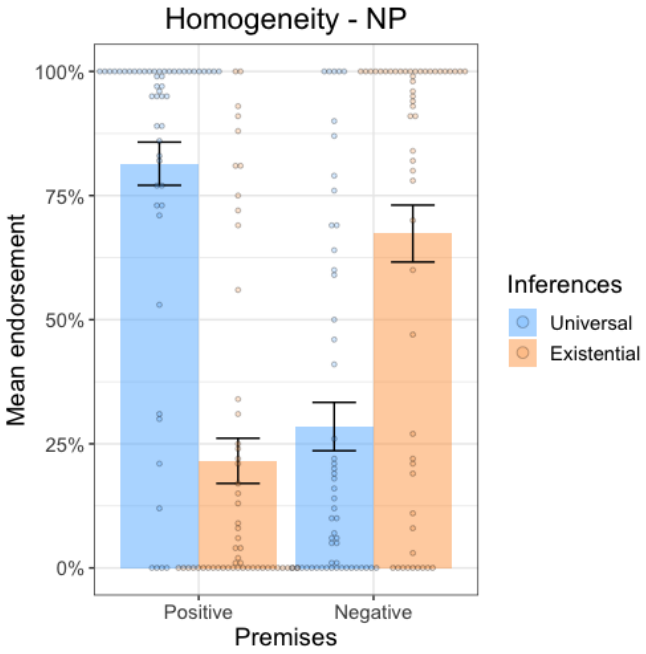

Figure 7: Reading flipping under negation

Iconicity controls

For each iconicity control, we were interested in the difference in endorsement between the inference containing the matching interpretation of the iconic modulation and the inference that contained the opposite non-matching interpretation. For instance, we wanted to know whether the inference containing the interpretation of a long upward scale as a high mountain was more endorsed than the inference containing the interpretation of a long upward scale as a low mountain. We thus counted inference type as a fixed effect. To model the possible differences in the tendencies to endorse both types of inferences in a certain way for each participant, we included inference type by subject as a random effect in the model, as shown:

Model

`value ~ InferenceTypeC + (1 + InferenceTypeC|SubjID)`

Control	(Intercept)	InferenceC
1	55.09	52.11
2	52.75	-28.37

Appendix III: Inferential mechanisms

Here, we provide a loose description of the underlying mechanisms responsible for the four types of inferences tested. Although an important part of its content is itself subject to debate, the aim of this table is merely to provide some very basic analysis of each inference to facilitate the reading of each section; it does *not* aim at exhaustively accounting for all possible theoretical models.

Inference	Input	Mechanism	Output
PRESUPPOSITION	Bivalent[22] truth-conditions (from either linguistic or non-linguistic items)	Takes the input q and produces a pair $<p, p'>$ with q equivalent to p & p' where p is the presupposition.	Presupposition + Assertion[23]
SCALAR IMPLICATURE	Alternatives (non-necessary linguistic) with their bivalent content.	Competition between alternatives (Gricean or grammatical mechanism)	Ordinary meaning + Strengthened meaning
SUPPLEMENT	Bivalent content + syntactic form	Whatever allows a non-restrictive relative clause to get the meaning of a supplement [e.g. comma intonation (Potts, 2015)]	Supplemental inference
HOMOGENEITY INFERENCE	*Previous theories:* Definite plural NP [or non-linguistic item referring to definite pluralities such as punctuated gestures or sounds] _____ *More recently* (Križ, 2020): Trivalent predicate [or non-linguistic predicate like gestural or musical]	Assigns the input a universal reading ('all') in positive environments and an existential reading in negative environments _____ If a group is in the extension of the predicate, then all members of the group are in the extension of the predicate as well (positive environment) or none of them (negative environments)	Homogeneous inference

Figure 8: Description of each inferential type

[22] Bivalence refers to the existence of pre-conditions to the meaning.

[23] The output is actually closer to <presupposition, presupposition+assertion>, because the assertion in <presupposition, assertion> is underdetermined.