

# Co-linguistic content inferences: From gestures to sound effects and emoji

Robert Pasternak  
Leibniz-Center for General Linguistics (ZAS)  
mail@robertpasternak.com

Lyn Tieu  
Western Sydney University  
lyn.tieu@gmail.com

October 27, 2021

---

Among other uses, co-speech gestures can contribute additional semantic content to the spoken utterances with which they coincide. A growing body of research is dedicated to understanding how inferences from gestures interact with logical operators in speech, including negation (“not”/“n’t”), modals (e.g., “might”), and quantifiers (e.g., “each”, “none”, “exactly one”). A related but less-addressed question is what kinds of meaningful content other than gestures can evince this same behavior; this is in turn connected to the much broader question of what properties of gestures are responsible for how they interact with logical operators. We present two experiments investigating sentences with *co-speech sound effects* and *co-text emoji* in lieu of gestures, revealing a remarkably similar inference pattern to that of co-speech gestures. The results suggest that gestural inferences do not behave the way they do because of any traits specific to gestures, and that the inference pattern extends to a much broader range of content.

**Keywords:** co-linguistic content, gesture, emoji, semantics, pragmatics

---

## I Introduction

*Co-speech gestures*—gestures temporally aligned with spoken content—serve a variety of purposes, including lightening the cognitive burden of language production (Krauss 1998; Goldin-Meadow et al. 2001; Gillespie et al. 2014); marking information-structural features like focus (Wilmes 2009; Ebert et al. 2011); and contributing additional meaningful content to the speaker’s utterance. An instance of the latter use of gestures can be seen in (1), in which the upward-pointing gesture UP coincides with the verb phrase “use the stairs”, generating an inference that the girl will take the stairs upward.

- (1) The girl will [use the stairs]<sub>UP</sub>.  
    ↪ The girl will go up the stairs.

Recent years have seen growing interest among formal semanticists in this last use of gestures, and in the precise nature of the meaningful contributions these gestures make (Lascarides & Stone 2009; Ebert & Ebert 2014; Davidson 2015; Schlenker 2018a; Tieu et al. 2017, 2018; Anvari 2017; Esipova 2018a,b, 2019; Zlogar & Davidson 2018; Hunter 2019). As will be discussed in the next section, much of this research has focused on the ways in which gestural inferences interact with logical operators in the spoken sentence, including **negation**

(“not”/“n’t”), **modals** (e.g., “might”), and **quantifiers** (e.g., “each”, “none”, “exactly one”). A related question, and one that has received considerably less attention in the literature to this point, concerns the extent to which whatever inference pattern is observed with gestures can also be observed with other kinds of meaningful content. This in turn relates to the much broader question of what precisely it is about co-speech gestures that makes their inferences behave as they do. Co-speech gestures are special in a variety of ways, including their important role in language development and their use of a visual modality distinct from the auditory modality of speech. The gestural inference pattern could conceivably be traced to any combination of these traits, leading to a fairly open space of plausible hypotheses, each of which makes different predictions about what kinds of content should behave in the same way (in relevant respects) as gestures.

In this paper we will address these questions by studying inferences from two other types of what we call *co-linguistic content*: *co-speech sound effects*, meaningful sound effects coinciding with spoken content; and *co-text emoji*, in which informative emoji—small images used in electronic communication—coincide with written text. Inferences from both kinds of content have been argued in the theoretical literature to interact with logical operators in the same way that gestures do (see Pasternak 2019 on sound effects, Pierini to appear on emoji), but until now these (often subtle) judgments have yet to be experimentally corroborated. With this in mind, we will report experimental evidence from inferential judgment tasks supporting the claim that inferences from co-speech sound effects and co-text emoji interact with logical operators in the same way that gestures do. The results of the sound effect and emoji studies bear a striking resemblance to those observed in a previous study using the same task to explore gestural inferences (Tieu et al. 2018). As a result we conclude that the apparent “gestural” inference pattern most likely encompasses a much broader array of content.

But first a brief note is in order regarding our use of the term *co-linguistic content*. By referring to co-speech gestures as “co-linguistic”, we do not necessarily wish to imply that gesture does not in some way interact with (or even constitute part of) the cognitive system commonly referred to as the *faculty of language*, i.e., the combinatorial apparatus that systematically generates complex expressions from discrete parts. We in fact take this to be an important open question that has yet to be fully resolved. But for the purposes of defining “co-linguistic content”, we intend “linguistic” to be construed extremely narrowly, encompassing only clear and obvious externalizations of discrete combinatorial syntactic processes—very roughly, pronounced or written strings of words. For the most part we will only be concerned with (auditory) speech and text, though we will briefly discuss sign languages as well. If gestures are indeed discovered to be partially or fully integrated parts of combinatorial linguistic cognition, then our findings remain of equal value, but their contribution takes on a slightly different hue: more specifically, they suggest that the manner in which gestures are integrated with speech in the compositional semantics extends well beyond gestures, or at least that other forms of inferential content like sound effects and emoji can exploit the same compositional mechanisms to derive analogous interpretations.

## 2 Background

### 2.1 Inferences and the operators they interact with

A growing body of work investigates how inferences from co-speech gestures interact with spoken logical operators. As an illustration of what we generally mean by “interaction with logical operators”, consider the gesture-less examples in (2).

- (2) a. Mary hired a tall architect.
- b. Mary likes the tall architect she hired.
- c. Mary hired an architect, who is tall.

There are superficial semantic parallels between the sentences in (2): each, for example, leads one to infer that Mary hired a tall person, and in particular a tall architect. However, the inferences derived from the three sentences vary in ways that can readily be observed through interactions with logical operators. Take, for example, sentential negation (“not”/“n’t”). The sentences in (3) are the result of introducing sentential negation into the sentences in (2):

- (3) a. Mary did not hire a tall architect.
- b. Mary does not like the tall architect she hired.
- c. # Mary did not hire an architect, who is tall.

The results of introducing negation differ substantially in the three cases, in spite of their superficial similarities in (2). Clearly, (3a) differs from (2a) in flatly contradicting the claim that Mary hired a tall architect, though whether Mary hired any short architects or tall non-architects is neither confirmed nor denied. In contrast, in (3b) we in fact retain the inference that Mary hired a tall architect (and, *a fortiori*, a tall person), in spite of the presence of sentential negation. Finally, in (3c), not only do we not retain an inference that Mary hired a tall architect, but the sentence as a whole becomes deviant (i.e., semantically or pragmatically ill-formed), as customarily indicated by a # sign.

We can further test similarities and differences through interactions with quantificational noun phrases like “exactly one of these three managers”, as in (4):

- (4) a. Exactly one of these three managers hired a tall architect.
- b. Exactly one of these three managers likes the tall architect she hired.
- c. Exactly one of these three managers hired an architect, who is tall.

What we infer from (4a) is that there is exactly one manager who hired a tall architect; other managers may have hired short architects or tall non-architects, but no others hired tall architects. Meanwhile, from (4b) we infer that *each* of the managers hired a tall architect, a so-called *universal* inference (Chemla 2009). And finally, from (4c) we infer that exactly one manager hired an architect *at all* (tall or otherwise), and that the architect that that manager hired is tall. Interactions with quantificational noun phrases thus further illustrate the contrasts between the inferences introduced in the sentences in (2).

While explanations of these inferential interactions involve complexities beyond the scope of this paper, the essential principle at play is that the way an inference will interact with logical operators like negation and quantifiers depends on how that inference is introduced. For example, the inference from (2a) to the conclusion that Mary hired a tall person is an *entailment*: it falls out straightforwardly from the “plain”, asserted (or *at-issue*) content of the sentence. Meanwhile, the inference in (2b) that Mary hired a tall person, and in particular a tall architect, stems from the referential noun phrase “the tall architect she hired”. Such referential noun phrases introduce *existential* inferences that a potential referent exists, e.g., that there is some tall architect that Mary hired. Moreover, this existential inference is not an entailment, but is *presupposed*: a sentence with a referential noun phrase that cannot successfully refer is generally not interpreted as false, but rather as deviant (Strawson 1950). Such presuppositions behave differently from entailments in a variety of ways, and there is a vast literature dedicated to determining and explaining precisely how presuppositions behave when embedded under operators like negation and quantifiers (so-called *presupposition projection* and *accommodation*).<sup>1</sup> Finally, in (2c) the inference that Mary hired a tall person is due to the presence of the non-restrictive relative clause *who is tall*; this falls under the broader category of *supplements*, which behave in a way that differs from both entailments and presuppositions (Potts 2005). It is worth noting that entailments, presuppositions, and supplements do not exhaust the class of kinds of inference derived in natural language; the point, rather, is that there are often several different ways one could generate a given inference in natural language, and how that inference will be affected by surrounding logical operators depends on the manner in which that inference was introduced in the first place.

## 2.2 Gestural inference interactions

In light of the observation that co-speech gestures can introduce additional inferences to the utterances with which they coincide, as well as the observation that how an inference interacts with logical operators depends on how that inference is introduced, a natural question—and one that has received increased attention in the recent literature—is how inferences derived from co-speech gestures interact with spoken logical operators. Analyses of the behavior of co-speech gestural inferences typically operate under the (by no means *a priori* obvious) assumption that this behavior is also observed in spoken language outside of gestures, with the choice of linguistic parallel determining ensuing empirical predictions.<sup>2</sup> However,

<sup>1</sup> For a sampling of important theoretical work on presuppositions, projection, and accommodation, see Strawson (1950); Karttunen (1973); Stalnaker (1973); Lewis (1979); Heim (1983, 1992); van der Sandt (1992); Geurts (1998); Schlenker (2009). For some experimental work, see Chemla (2009); Chemla & Schlenker (2012); Chemla & Bott (2013); Domaneschi et al. (2013); Schwarz (2007, 2015, 2019); Schwarz & Tiemann (2017).

<sup>2</sup> For example, Ebert & Ebert (2014) argue that gestures contribute supplements; Schlenker (2018a) argues that gestures contribute a special kind of presupposition that he calls a *cosupposition*; Esipova (2019) argues that gestures can be compositionally integrated in all and only those ways independently attested in spoken language, with differences from spoken language being accounted for through pragmatic principles; and Hunter (2019) argues for an approach in which many gesture-induced inferences are accounted for through indepen-

for our purposes theoretical analyses attempting to explain the behavior of co-speech gestural inferences are of relatively little import, as our concern is focused instead on what kinds of content give rise to this behavior. With this in mind, here we will discuss the intuitive judgments offered by Schlenker (2018a), as these are the ones tested by Tieu et al. (2017, 2018); later we will discuss Tieu et al.'s (2018) inferential judgment task study confirming that Schlenker's gesture judgments are shared by untrained speakers, in addition to our own study of sound effects and emoji.

A seemingly plausible hypothesis about the semantic contribution of UP in (1) is that it simply serves as an at-issue verb phrase modifier, akin to an adverbial prepositional phrase like “in an upward direction” or “[in this direction]<sub>UP</sub>” (where deictic “this” explicitly refers to the semantic content of the gesture):

- (5) a. The girl will use the stairs in an upward direction.  
 b. The girl will use the stairs [in this direction]<sub>UP</sub>.

We will refer to examples like (5) as including *at-issue modification*, or simply as *at-issue examples*. For the sake of convenience, we will reserve the term *co-speech gestures* for cases like (1), where the gesture makes its semantic contribution without being explicitly incorporated into the content of an at-issue modifier as in “[in this direction]<sub>UP</sub>”.

In accordance with our simple hypothesis, (5a) and (5b) do in fact seem to have roughly the same interpretation as (1): all are true if and only if the girl will go up the stairs. However, as was shown in examples (2) through (4), inferences that look identical in unembedded environments can diverge considerably once logical operators are introduced. Schlenker (2018a) argues that such divergence can indeed be observed in a variety of cases. Take, for instance, sentential negation, illustrated with the at-issue examples in (6):

- (6) a. The girl will not use the stairs in an upward direction.  
 b. The girl will not use the stairs [in this direction]<sub>UP</sub>.

The truth conditions for the sentences in (6) are straightforward: the girl will not go up the stairs. Importantly, the truth of these sentences is compatible with the girl either not taking the stairs at all, or going down the stairs; all that is required is that she not go upward. Now consider (7):

- (7) The girl will not [use the stairs]<sub>UP</sub>.  
 ~> If the girl were to use the stairs, she would go up the stairs.

An utterance of (7) appears to carry stronger truth conditions than (6). First, it requires that the girl not take the stairs *at all*, including downward; in other words, the part of the utterance that excludes the gesture (“The girl will not use the stairs”) must be true. And second, it introduces a *conditional inference* that if the girl *were* to use the stairs, she would go upward; this conditional inference is not derived in (6). We thus see that while examples like (1) and (5) appear to be semantically identical, once negation is introduced differences are immediately observable between co-speech gestures and at-issue modification.

---

dently attested discourse relations.

This divergence can be further observed in interactions with the modal verb “might”. Consider the at-issue examples in (8):

- (8) a. The girl might use the stairs in an upward direction.  
 b. The girl might use the stairs [in this direction]<sub>UP</sub>.

The claims made by (8a) and (8b) are again straightforward: each is true if there is some possibility of the girl going up the stairs, even if it is equally plausible that the girl will instead go down the stairs. This is not the case with (9):

- (9) The girl might [use the stairs]<sub>UP</sub>.  
 ~> If the girl were to use the stairs, she would go up the stairs.

(9), like (7), generates two salient inferences: the truth of the gesture-less portion of the utterance (“The girl might use the stairs”) and a conditional inference that if the girl uses the stairs, it will be to go upwards. As a result, (9) differs from (8) in that it permits the possibility that the girl will take the stairs upwards, or that she will not take the stairs at all, but *not* that she will take the stairs downward.

Finally, at-issue modification and co-speech gestures can be further teased apart through their interactions with quantifiers like “each”, “none”, and “exactly one”. Consider the co-speech gesture examples in (10):

- (10) a. Each of these three girls will [use the stairs]<sub>UP</sub>.  
 b. None of these three girls will [use the stairs]<sub>UP</sub>.  
 c. Exactly one of these three girls will [use the stairs]<sub>UP</sub>.

~> For each of these three girls, if she were to use the stairs, she would go up the stairs.

Much like in the previous cases, each of the sentences in (10) requires that the sentence absent the gesture be true: for (10a) to be true each girl must use the stairs, for (10b) none must use the stairs (in any direction), and for (10c) exactly one must use the stairs. In addition, each sentence introduces a *universal conditional inference*: it must be the case that for *each* of the three girls, if she *were* to use the stairs, it would be in an upward direction.

The examples in (10) can be compared and contrasted with their at-issue modification counterparts in (11); these examples feature “[in this direction]<sub>UP</sub>”, but parallel observations can be made about examples with “in an upward direction”.

- (11) a. Each of these three girls will use the stairs [in this direction]<sub>UP</sub>.  
 b. None of these three girls will use the stairs [in this direction]<sub>UP</sub>.  
 c. Exactly one of these three girls will use the stairs [in this direction]<sub>UP</sub>.

(11b) and (11c) make very different claims from their co-speech gesture counterparts (10b) and (10c). (10b) requires that none of the girls use the stairs at all, and that each would go up if she did use the stairs; (11b) merely requires that none of the three girls will go up the stairs, leaving open the possibility that some or all of them will go down the stairs. Similarly, (10c) requires that exactly one girl use the stairs at all, and that each girl would go up if



she did use the stairs; (11c) only requires that exactly one girl go *up* the stairs, so that the other two can either not take the stairs at all or take the stairs downward. Note, however, that (10a) and (11a) do seem to have identical interpretations: both require that each of the three girls go up the stairs. If (10a) behaves like (10b) and (10c) in requiring the truth of (i) the gesture-less part of the sentence (“Each of these three girls will use the stairs”) and (ii) the universal conditional inference that each girl would go up if she used the stairs, this observation is unsurprising: after all, by conjoining these inferences what we get is precisely that each girl will take the stairs upward. In other words, (12a) ( $\approx$  (10a)) and (12b) ( $\approx$  (11a)) are truth-conditionally identical:

- (12) a. Each girl will use the stairs, and each girl will go up if she uses the stairs.  
 b. Each girl will go up the stairs.

To summarize, here are the judgments as reported by Schlenker (2018a). Sentences with co-speech gestures appear to generate two (relevant) inferences. The first inference, presumably due specifically to the spoken content of the utterance, is simply that the gesture-less portion of the sentence is true on its own. The second inference, and the one most directly attributable to the co-speech gesture itself, is a plain modifying inference in unembedded examples like (1) (“The girl will go up the stairs”); a conditional inference in examples with negation and “might” like (7) and (9), respectively (“If the girl were to use the stairs, she would go up the stairs”); and a universal conditional inference in quantificational examples like (10) (“For each of these three girls, if she were to use the stairs, she would go up the stairs”). We also saw that in unembedded examples like (1) and examples with “each” like (10a), these ended up being equivalent to parallel examples with at-issue modification like “in an upward direction” and “[in this direction]<sub>UP</sub>”, but that this was not the case in other environments. Tieu et al. (2018) provide experimental evidence that untrained speakers share Schlenker’s judgments about the gesture-derived inferences in these examples and their contrast with parallel at-issue examples; we will return to their study when we discuss our own experiments with sound effects and emoji.

### 2.3 Possible sources of the inference pattern: The view from the theoretical literature

While we will soon see experimental evidence from Tieu et al. (2018) suggesting that Schlenker’s (2018a) judgments are shared by non-linguist speakers, for now let us simply assume that Schlenker’s characterization of the inferences in these examples is accurate. An important and natural follow-up question that has not been satisfactorily answered in the literature thus far is what traits of co-speech gestures are responsible for their inferences behaving in the particular fashion that they do. Or, put another way, do any other kinds of secondary content give rise to inferences that behave in the same way as gestures do? If so then which, and why? And if not, then why not?

One could formulate a large number of seemingly plausible hypotheses about what kinds of secondary content give rise to the same inference pattern as co-speech gestures. Here we

will discuss a small sample of what seem to us to be in-and-of-themselves reasonable hypotheses based on certain natural classes that co-speech gestures fall into, with the understanding that there may be numerous other, equally reasonable hypotheses. We will then discuss what the theoretical literature has had to say about the hypothesis space; while the questions at hand have not received much in the way of direct discussion, various other kinds of content have been argued to behave in the same way as co-speech gestures, and in each case the truth or falsehood of the empirical claims would have substantial effects on which hypotheses remain viable and which are likely false.

Presumably, the most restrictive viable claim in the hypothesis space would be **co-speech gesture uniqueness**: the inference pattern observed by Schlenker is restricted only to co-speech gestures.<sup>3</sup> Gestures are an important part of human communication in general, and linguistic communication in particular. We gesture constantly when we are speaking, and often when we are not. In addition to marking information structure and facilitating production, as mentioned in the introduction, gestures also play an important role in language acquisition (Iverson & Goldin-Meadow 2005), and are so deeply embedded in human language cognition that not only do congenitally blind children gesture when they speak, but they even gesture when speaking with other children that they know to be blind (Iverson & Goldin-Meadow 1997, 1998). Furthermore, non-human primates have been shown to make rich use of communicative gestures (Byrne et al. 2017), suggesting that gesture may have played an important evolutionary role in interhuman communication. The gestural inference pattern may thus be tied to some combination of traits that is unique to gestures, such as their early and frequent use in acquisition and communication, or perhaps some genetic component arising during the course of human evolution.

Another, less restrictive hypothesis about the origins of the co-speech gestural inference pattern would be that it stems from the **multimodality** of speech-gesture pairings. Gestures are obviously visual (and perhaps tactile) stimuli, while spoken utterances are equally obviously auditory, at least for non-signed languages. As a result, when a speaker produces an utterance containing both spoken and gestural content, their interlocutors are simultaneously interpreting information from two distinct streams of sensory input. It is possible that the gestural inference pattern is a result of such multimodality of speech and gesture. If this is the case, we might expect the relevant inference profile to be exhibited not only by gestures, but also by other types of visual co-speech content like pictures, animations, or signage.

A third possibility is that the co-speech inference pattern observed above is a feature of **embodied content**, i.e., co-speech content that is in some sense directly produced by the

<sup>3</sup> Note that “most restrictive” does not mean the same thing as “strongest”. Whether a hypothesis is *more restrictive* than another is a matter of which hypothesis predicts the inference pattern to arise with the smallest amount of content. Whether a hypothesis is *stronger* than another is a matter of entailment: a hypothesis  $\varphi$  is stronger than hypothesis  $\psi$  if whenever  $\varphi$  is true,  $\psi$  must also be true (but not vice versa). The hypotheses listed in the body of the paper vary in their restrictiveness, but not in their strength, since all are mutually contradictory: for example, the hypothesis that the inference pattern is a general phenomenon of co-linguistic content is not weaker than the claim that it is restricted to gestures, since the two claims cannot both be true.



speaker's body.<sup>4</sup> Unlike the multimodality hypothesis, this would exclude visual co-speech information not (perceived as being) produced by the speaker's body (e.g., animations), but it would include embodied content that occupies the same modality as speech, such as vocal modulations and clapping noises.<sup>5</sup>

A less restrictive hypothesis than the first three would be to posit that the gestural inference pattern is observed across all forms of **co-linguistic content**. That is, co-speech gestures do not behave as they do because of any traits that are specific to gestures, but simply because they are co-linguistic content in the first place. According to this hypothesis, all else being equal we expect the gestural inference pattern to arise for a wide variety of co-linguistic content, including those kinds mentioned above, as well as sound effects and emoji, to be discussed later.

Finally, perhaps the least restrictive hypothesis would be that the inference pattern at hand is not even restricted to *co-linguistic* content: any instance in which some secondary meaningful content coincides with some primary meaningful content will give rise to the same inference pattern. For example, Schlenker (2018a) observes that co-speech uses are not the only ways in which gestures can be integrated with speech: there are also *pro-speech* uses in which the gesture plays the role of some normally spoken constituent, rather than coinciding with one:

(13) The girl will UP. (≈ 'The girl will go upwards.')

If one aligns a second gesture—perhaps a non-manual gesture, as suggested by Esipova (2019)—with the pro-speech gesture UP, the **secondary content** hypothesis would predict that the inference from the secondary gesture should behave just like co-speech gestures, while the **co-linguistic content** hypothesis would predict that it should not (all else being equal), since the secondary gesture is not aligned with linguistic material (on the extremely narrow definition of "linguistic" discussed in the introduction).<sup>6</sup>

<sup>4</sup> Note that the embodied content hypothesis is about the *manner* in which conceptual content is conveyed, rather than the concepts themselves: an animation of someone kicking a ball does not count as embodied content, even though the concept of ball-kicking could be considered an embodied concept.

<sup>5</sup> This is not to say embodied content *must* occupy the same modality as speech: co-speech gestures, for example, are also embodied content. We bring up same-modality embodied content only to draw a distinction between the multimodality and embodied content hypotheses.

<sup>6</sup> Esipova (2019) discusses such cases of what might be called "co-gesture gestures". The primary gesture in Esipova's examples is a Russian gesture DRUNK, which she takes to mean "'drink (alcohol)' or 'drunk'". DRUNK has two variants, one in which the speaker flicks her finger on her neck, and another in which she taps her neck with the back of her hand. The secondary, non-manual gesture is what Esipova calls O\_O, a wide-eyed look of surprise. Thus, she considers examples like (i):

(i) Yesterday there was a party, and Mia got [DRUNK]<sub>O\_O</sub>.  
(≈ 'Mia got surprisingly drunk.')

Esipova does not directly address how inferences from O\_O interact with the particular operators in question, and the specific choices of examples involve complexities well beyond the scope of this paper. Co-gesture gestures are a fascinating and underexplored issue that we leave for future work, though see fn. 7 for discussion of potential hurdles facing empirical investigation of this phenomenon.

So what does the theoretical literature suggest about the viability of these various hypotheses? Put succinctly, arguments put forward in the theoretical literature tend to favor less restrictive hypotheses like **co-linguistic content** and **secondary content** over more restrictive hypotheses like **co-speech gesture uniqueness**. A brief summary of some examples can be seen below.

**Iconic vocal modulations** Schlenker (2018b) discusses an example attributed to Robert Pasternak (pers. comm.) in which iconic vocal modulation introduces inferences:

- (14) Johnny is about to start [talking]<sup>high-pitched</sup>.  
 ~> Johnny will speak in a high-pitched voice.

In an example like (14), saying “talking” in a high-pitched voice can lead to an inference that any speech by Johnny will be high-pitched. Moreover, when embedded under negation we retain a conditional inference about Johnny’s speech, much like with gestures:

- (15) Johnny is not about to start [talking]<sup>high-pitched</sup>.  
 ~> If Johnny were to speak, it would be in a high-pitched voice.

Examples like these suggest that inferences from vocal modulations might behave in the same fashion as gestures. This would be problematic for **co-speech gesture uniqueness** and **multimodality** (since iconic vocal modulations occur in the same auditory modality as speech itself), but not for the other hypotheses.

**Co-sign gestures** In a similar vein, Emmorey (1999), Davidson (2015), Aristodemo (2017), Goldin-Meadow & Brentari (2017), and Schlenker (2018b) argue in favor of the existence of “co-sign” gestures, i.e., gestures coinciding with signs in signed languages. Aristodemo (2017) and Schlenker (2018b) in particular argue that not only do co-sign gestures exist in Italian Sign Language (LIS) and American Sign Language (ASL), respectively, but they interact with logical operators in the same way as their co-speech counterparts. Since these would also involve cases of gesture and sign occupying the same modality, these empirical claims would be problematic for **multimodality**, and perhaps for **co-speech gesture uniqueness**, depending on if co-speech and co-sign gestures should be treated qualitatively differently or as the same in relevant respects.

**Co-speech sound effects** Pasternak (2019) discusses cases of *co-speech sound effects*, which as the name suggests are meaningful sound effects coinciding with speech. This is illustrated in Pasternak’s (16), in which the explosion sound *explode* is aligned with the verb phrase “assassinate his target”, generating an inference that the assassination will be by means of an explosion. (Pasternak’s sound effect examples can be found at <https://bit.ly/2Je6Sto>.)

- (16) The soldier will [assassinate his target]<sub>explode</sub>.

Co-speech sound effects lack many of the traits that make gestures so special. By all appearances they did not play any significant role in language evolution, nor are they important

to language acquisition. Outside of the seemingly narrow purview of radio programs, podcasts, and audiobooks, they do not frequently appear in day-to-day language use. Speech-sound pairings are not multimodal—each uses the auditory modality—and many sound effects, including those used in the study reported in this paper, could not plausibly be directly produced by the human body. Nonetheless, Pasternak argues that co-speech sound effects do in fact interact with logical operators in the same way gestures do, citing as evidence speech-sound pairings like the negated (17), which generates a conditional inference much like (7) does:

- (17) The soldier will not [assassinate his target]<sub>explode</sub>.  
 ↪ If the soldier were to assassinate his target, he would do so via explosion.

If Pasternak's (2019) observations about co-speech sound effects are correct, then they are problematic for all of the accounts suggested above other than the **co-linguistic content** and **secondary content** hypotheses, since co-speech sound effects are obviously not co-speech gestures, nor are they multimodal or embodied content.

**Co-text emoji** Yet another case, discussed in detail by Pierini (to appear), are *co-text emoji*. Emoji are small images encoded as text, often integrated with writing on social media and in digital messaging. On separate grounds, emoji have previously been argued to serve a communicative function similar to gestures (Gawne & McCulloch 2019). With respect to interactions with logical operators, Pierini (to appear) argues that emoji behave like gestures in a variety of environments, including co-text emoji as a particular sub-case. In illustrating Pierini's point, we will modify the presentation of examples somewhat: whereas Pierini's examples involve emoji simply following the text of the sentence, as in (18), in our examples both in this paper and in the reported experiment, emoji "bracket" the verb phrase on either side, as in (19):

- (18) The student will step out of the classroom 🚽  
 (19)           The student will  
           🚽 step out of the classroom 🚽

While the former comes across as more natural, our reason for using the latter is that, as Schlenker (2018a) and Esipova (2019) discuss in detail, *post-speech* gestures—gestures following and not coinciding with speech—behave differently from *co-speech* gestures. Pierini makes similar observations about emoji. Thus, putting the emoji at the end of the sentence could lead to an ambiguity between a "co-text" and "post-text" interpretation, a problem that emoji bracketing seems to us to avoid. That being said, we also piloted the same experiment without the bracketing format and found similar results.

(19) generates an inference that the student will step out of the classroom in order to use the toilet. But Pierini notes that when embedded under a variety of operators, the gesture inference pattern seems to arise with emoji too. For instance, (20) with sentential negation gives rise to a conditional inference parallel to the gestural inference in (7) and the sound effect inference in (17):

- (20) The student will not  
 🙅 step out of the classroom 🙅

↪ If the student were to step out of the classroom, it would be to use the toilet.

Pierini provides similar examples with “might”, “each”, “exactly one”, and “none”, arguing that in each case, emoji-derived inferences behave like gestures. If Pierini’s observations are correct, then emoji are like sound effects in being problematic for the **co-speech gesture uniqueness, embodied content**, and presumably **multimodality** accounts, on the assumption that text and emoji count as using the same (visual) modality.

All of these empirical observations, if valid, constrain the hypothesis space in a way that generally favors less restrictive over more restrictive hypotheses, since more content gives rise to the gestural inference pattern than the more restrictive hypotheses would predict. However, all of the empirical observations discussed above are reports of intuitive judgments by individual speakers. While individual judgments frequently suffice for the purposes of linguistic research, judgments pertaining to the semantics and pragmatics of co-linguistic gestures, sound effects, emoji, vocal modulations, etc. are often subtle and unstable, especially when it comes to how these inferences interact with logical operators. This was a major motivating factor for Tieu et al.’s (2017, 2018) experimental studies on gestures: because of the subtlety and instability of judgments, Tieu et al. sought to experimentally confirm that non-linguist speakers share the same judgments as those reported in Schlenker’s (2018a) theoretical work. Since non-gesture inferences are often at least as subtle and unstable as their gestural counterparts, it is important to experimentally confirm that the gestural inference pattern indeed extends to the other kinds of content that have been suggested in the literature. In Section 3 we discuss two experiments testing precisely this for co-speech sound effects and co-text emoji. The results of the two experiments replicate those of the gestural experiments reported in Tieu et al. (2018), supporting claims by Pasternak (2019) and Pierini (to appear) that co-speech sound effects and co-text emoji behave like gestures.

While we believe that all of the observations discussed above warrant thorough experimental corroboration, there are a couple of reasons why co-speech sound effects and co-text emoji make for an especially convenient place to start in testing what types of content interact with logical operators in a gesture-like fashion. First, sound effects and emoji are very efficient in terms of eliminating possible hypotheses from the hypothesis space. As discussed above, for both co-speech sound effects and co-text emoji, if inferences are found to behave in a gesture-like manner then this would serve as evidence against the **co-speech gesture uniqueness** hypothesis (since neither are gestures), the **multimodality** hypothesis (since in both cases the linguistic and co-linguistic content occupy the same modality), and the **embodied content** hypothesis (since in both cases the co-linguistic content is not perceived as coming from the speaker’s body). Similar considerations likely extend to many other plausible hypotheses not considered here: there are many traits that co-linguistic sound effects and emoji do not share with gestures, and thus many traits that can be eliminated as explanations for the gestural inference profile. And second, co-speech sound effects and co-text

emoji allow for a very clean differentiation between primary content (speech or text) and secondary content (sound effect or emoji). In other cases things can be less clear. Take, for example, co-sign gestures. Since gestures and signs occupy the same modality and use the same articulators, determining which parts of a given signed utterance are sign vs. gesture is a non-trivial matter: what one person calls a “co-sign gesture” another person might simply call “part of the sign”. Thus, until co-sign gestures can be confirmed as being such, the theoretical implications for the behavior of their inferences are less clear than in cases like co-speech sound effects and co-text emoji, where there is a clean break between linguistic and co-linguistic content.<sup>7</sup>

With this in mind, in the next section we discuss two experiments replicating Tieu et al.’s (2018) gesture results with co-speech sound effects and co-text emoji. The success of this replication allows us to significantly trim down the space of hypotheses, favoring **co-linguistic content** and **secondary content** over **co-speech gesture uniqueness**, **multi-modality**, **embodied content**, and other similarly restrictive hypotheses.

### 3 Experiments

#### 3.1 Tieu et al.’s gesture study

Tieu et al. (2018) use an inferential judgment task (IJT) to test the behavior of gestural inferences in the six environments discussed above: UNEMBEDDED (cf. (1)), MIGHT (cf. (9)), NEGATION (cf. (7)), and the quantificational environments EACH (cf. (10a)), NONE (cf. (10b)), and EXACTLY ONE (cf. (10c)). On each trial, participants were presented with a video of one of the experimenters uttering a sentence containing a gesture; text was provided below the video indicating a particular potential inference, and participants were asked to indicate with a slider scale how strongly they derived that inference, ranging from “Not at all” to “Very strongly”. The scale was otherwise unmarked, but each input was mapped to an integer from zero to 100.

For the non-quantificational environments (UNEMBEDDED, MIGHT, NEGATION), the inference tested was the one provided in the examples in this paper ((1), (9), and (7), respectively), i.e., a simple “will go up the stairs” inference for UNEMBEDDED, and conditional inferences for MIGHT and NEGATION. Whereas only one inference was tested in these non-quantificational environments, for each quantificational environment (EACH, NONE, EXACTLY ONE) two inferences were tested in separate trials: **existential** and **universal** conditional inferences, illustrated in (21):

(21) a. **Existential:** For at least one of these three girls, if she were to use the stairs, she

<sup>7</sup> Tests of co-gesture gestures like in fn. 6 face similar practical hurdles. First, in order for tests of co-gesture gestures to be meaningful with respect to the hypothesis space discussed above, one must be able to demonstrate that the test examples are being perceived as having two simultaneous gestures (DRUNK and O\_O), rather than a single, complex gesture (DRUNK + O\_O). Moreover, one must also demonstrate that one gesture is interpreted as primary and the other as secondary. These are not insurmountable hurdles by any stretch, but we leave the resolution of these issues for future work.

would go up the stairs.

- b. **Universal:** For each of these three girls, if she were to use the stairs, she would go up the stairs.

Notice that the universal inference asymmetrically entails the existential one: if something holds of all three girls, it also holds of at least one, but not vice versa. Thus, participants who endorse a universal inference for a given sentence are expected to also endorse an existential inference for that sentence.

For each pairing of environment and inference, control items were also included in which the verb phrase “[use the stairs]<sub>UP</sub>” was replaced with the at-issue “use the stairs [in this direction]<sub>UP</sub>”. **Condition (TARGET vs. CONTROL)** was a between-subject factor: one group of participants saw only target items, and one group only control items. As per the discussion above, in the case of both UNEMBEDDED and EACH, endorsement rates were predicted to be high for both target and control items, since in these environments target and control items came out to the same interpretation. This includes both existential and universal conditional inferences for EACH: we saw above that EACH generates universal inferences in both target and control conditions, and existential inferences are strictly weaker than universal inferences. **For UNEMBEDDED and EACH then, no effect of condition is expected.** In the NEGATION, MIGHT, and NONE environments, the prediction is a significant **effect of condition**, with target items receiving higher inference endorsement rates than control items. This is true of both existential and universal inferences in the NONE environment: as illustrated in (22) below, neither inference is derived in the control condition, while we saw above that the universal (and thus also existential) inference is derived in the target condition.

- (22) None of these three girls will use the stairs [in this direction]<sub>UP</sub>.  
 ↗ For at least one of these three girls, if she were to use the stairs, she would go up the stairs.  
 ↗ For each of these three girls, if she were to use the stairs, she would go up the stairs.

This just leaves EXACTLY ONE. As we saw above, EXACTLY ONE does not generate a universal conditional inference in the control condition. However, it does entail an existential inference:

- (23) Exactly one of these three girls will use the stairs [in this direction]<sub>UP</sub>.  
 ↗ For at least one of these three girls, if she were to use the stairs, she would go up the stairs.  
 ↗ For each of these three girls, if she were to use the stairs, she would go up the stairs.

Thus, in the case of EXACTLY ONE we **expect an effect of condition for the universal inferences but not for the existential inferences: for the universal inferences, target items should receive significantly higher endorsement rates than control items, while for the existential inferences, endorsement rates should be high for both target and control items.**

To summarize, then, **for UNEMBEDDED, both existential and universal inferences for EACH, and existential inferences for EXACTLY ONE, no effect of condition was expected, with**






high endorsement rates predicted for both targets and controls; in all other cases, a significant effect of condition was expected, with target items predicted to elicit higher endorsement rates than control items.

The results of Tieu et al.'s study can be seen in Figure 3, where red bars (“Gesture”) indicate mean endorsement for target items, and blue bars (“Asserted”) indicate mean endorsement for controls. We will return to Tieu et al.'s results when we discuss those of our own experiments on co-speech sound effects and co-text emoji.

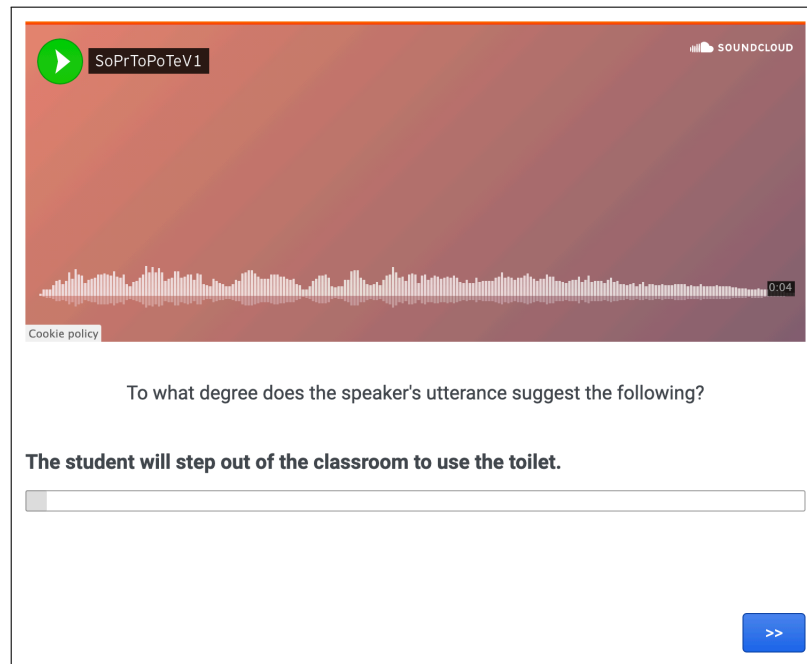
### 3.2 Sound effects and emoji

For the sound effect and emoji experiments, the materials, raw data, and R scripts for analysis can be accessed at <https://osf.io/5vh7m>. The experiments were advertised on Amazon Mechanical Turk (MTurk) as 15-minute studies at a pay rate of 3USD; each took on average 14 minutes to complete. Informed consent was obtained from all participants. Participants were directed to the web-based experiment, implemented and hosted on the Qualtrics platform. The experiments consisted of IJTs modelled after that of Tieu et al. (2018): on each trial, participants either listened to an embedded SoundCloud audio file of a sentence containing a sound effect (see Figure 1) or read a sentence containing an emoji (see Figure 2), and had to indicate the degree to which they accessed the inference indicated in text below the test item. Trials were self-paced: on each page, for the entire duration of the trial, participants could see the embedded SoundCloud waveform or image of the emoji-containing sentence, the inference to be judged beneath it, and a slider scale beneath that, which participants could use to indicate the degree to which they accessed the inference. The slider scale was unmarked to participants, but as in Tieu et al.'s experiment, was underlyingly mapped to a scale from 0–100 for data analysis.

Both experiments included target and control conditions, constructed as sound effect and emoji equivalents of the gestural targets and controls of Tieu et al. (2018) ((24) and (25), respectively). Tested inferences were also constructed in a parallel fashion, as in (26).

- (24) a. **Target:** The student will not [step out of the classroom]<sub>flush</sub>.  
 b. **Control:** The student will not step out of the classroom to do this: flush.
- (25) a. **Target:** The student will not  
 step out of the classroom   
 b. **Control:** The student will not step out of the classroom to do this: 
- (26) If the student were to step out of the classroom, it would be to use the toilet.

We tested various sound effects and emoji in the same six environments as in the gesture study: UNEMBEDDED, MIGHT, NEGATION, EACH, NONE, and EXACTLY ONE. As in the gesture study, the targeted inferences that we tested involved conditional inferences in the MIGHT and NEGATION conditions, and both existential and universal inferences in the quantificational cases. [The predicted patterns of endorsement for the target and control inferences in](#)

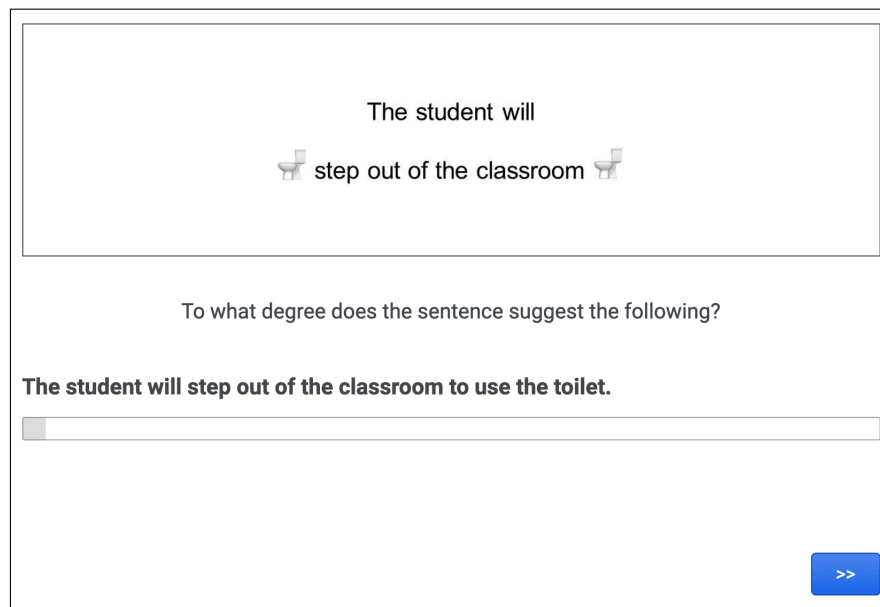


**Figure 1** Screen capture of a sound effect trial.

these various environments were the same as those described for the gesture study: for UN-EMBEDDED, both existential and universal inferences for EACH, and existential inferences for EXACTLY ONE, no effect of condition was expected, with high endorsement rates predicted for both targets and controls; in all other cases, a significant effect of condition was expected, with target items predicted to elicit higher endorsement rates than control items.

In the sound effect experiment, we tested five distinct sound effects: the opening of a can containing a carbonated beverage, a mirror breaking, a phone ringing, a plane taking off, and a toilet flushing. As indicated by the bracketing in (24), for target items, the sound effect aligned with the entirety of the verb phrase (starting at the onset of the verb phrase and lasting roughly until the end of the verb phrase), while for the controls the sound effect followed the deictic “this”. In the emoji experiment, we tested five distinct emoji: cigarette (🚬), rain-on-umbrella (☔), phone (📞), plane (✈️), and toilet (🚽).<sup>8</sup> To ensure that the

<sup>8</sup> A reviewer notes that the purported semantic relation between the linguistic and co-linguistic content seems to vary: for example, the contribution of the 🚽 emoji in the tested sentences involved a purpose relation (“The student will step out of the classroom to use the toilet”), while ✈️ contributed a manner reading (“The businesswoman will travel to the board meeting by plane”) and ☔ contributed a causal reading (“The party will be cancelled tomorrow because of rain”). We decided to test a variety of sound effects and emoji, to ensure that any observed effects were not an idiosyncrasy of a specific sound effect/emoji or class of sound effects/emoji. To take into account potential variability across sound effects/emoji, in analyzing the data, the mixed effect linear regression models we ran included random intercepts for sound effect/emoji, and thus we can be reassured that any observed main effects hold in spite of any variability across the sound effects/emoji. We also looked more closely at the results by sound effect/emoji and none appeared to differ particularly from



**Figure 2** Screen capture of an emoji trial.

emoji symbols displayed uniformly to all participants, what participants saw on the screen were actually screen captures of the sentences with their associated emoji. All sound effects and emoji were tested in all six linguistic environments.<sup>9</sup>

In each experiment, 200 MTurk workers were randomly assigned to either the target or control condition, for a total of 100 participants per condition. The sample size was judged to be sufficient given Tieu et al. (2018) had observed effects with 125 participants across the target and control conditions. With five sound effect/emoji types, six environments, and both existential and universal inferences in the quantificational cases, each participant judged a total of 45 sentence-inference pairs.

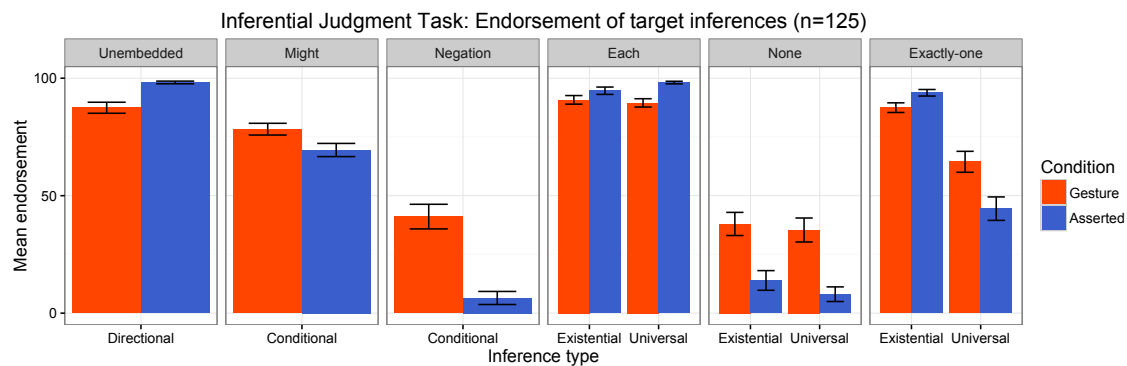
### 3.3 Results: Sound effects and emoji behave like gestures

The raw data and R scripts for analysis (including details of statistical models) can be accessed at <https://osf.io/5vh7m>. Figure 4 and Figure 5 show very similar endorsement rates to those for gestures in Tieu et al.'s Figure 3. Impressionistically, across the three types of

the others.

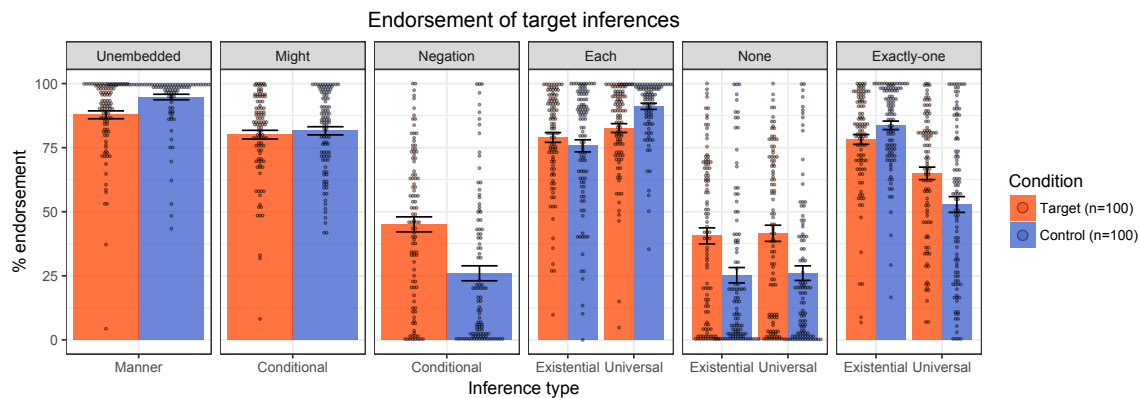
<sup>9</sup> We attempted to select highly iconic sound effects and emoji whose meanings would be relatively unambiguous. Although there was some overlap between the meanings depicted by the sound effects and the emoji (e.g., AIRPLANE, TOILET), we decided to prioritize clarity within each modality rather than a complete overlap across modalities. While we didn't ultimately observe any noteworthy differences across the sound effects or emoji that we tested, a future study could aim to more systematically test semantic equivalents across the different modalities (e.g., a manual gesture mimicking an explosion, a sound effect of an explosion, and the exploding bomb emoji). As an anonymous reviewer points out, it is possible that the kinds of inferences that people most naturally draw from these co-linguistic objects might differ depending on the modality.

co-linguistic content, there is strong endorsement of certain inferences, in particular those associated with UNEMBEDDED, MIGHT, the existential and universal inferences associated with EACH, and the existential inference associated with EXACTLY ONE. Following Tieu et al., we further wanted to know whether the tested inferences were *more strongly* endorsed for the sound effect and emoji target sentences than for their associated controls, as this would indicate that the co-linguistic content behaved differently from mere at-issue modification. To determine whether Condition (TARGET vs. CONTROL) was a significant predictor of the inferential judgment responses, we fitted mixed effect linear regression models to the data (using the lme4 package in R, R Core Team 2016; Bates et al. 2015), with Condition (TARGET vs. CONTROL) as a fixed effect, and random intercepts for participant and for type of emoji/sound effect. (The data for the sound effect and emoji experiments were analyzed separately, so there were 12 regression models in all, corresponding to each of the six linguistic environments in each of the two modalities.) We then used  $\chi^2$  statistics with one degree of freedom to compare models with and without the fixed effect of Condition. For the quantified conditions (EACH, NONE, and EXACTLY ONE), in addition to seeing whether target inferences were more strongly endorsed than control inferences, we also wanted to measure whether the strength of the effect was greater for one reading compared to the other (existential vs. universal); the models thus included as fixed effects Condition (TARGET vs. CONTROL), Reading (EXISTENTIAL vs. UNIVERSAL), and their interaction; random effects corresponded to by-participant slopes for Reading, since Reading was a within-subject factor.

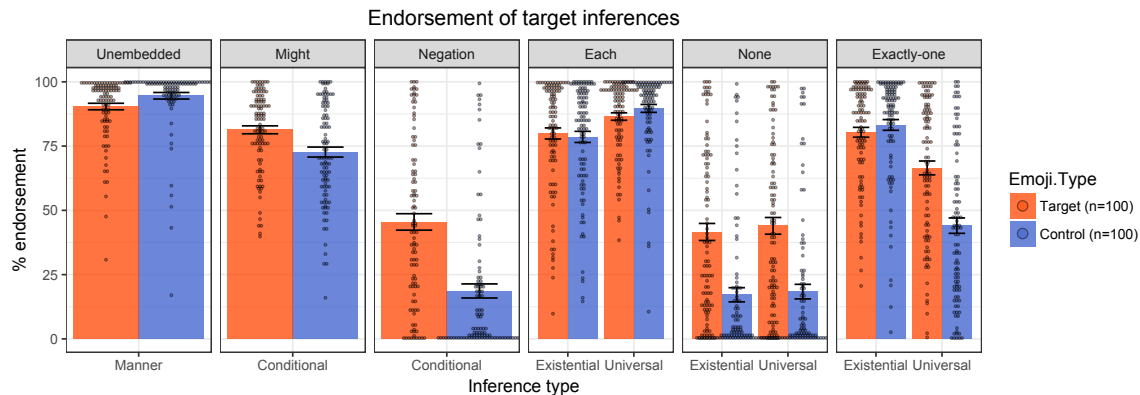


**Figure 3** Results of Tieu et al.'s (2018) inferential judgment task with gestures.

**Non-quantified environments: UNEMBEDDED, MIGHT, NEGATION** In the UNEMBEDDED environment, all three studies saw high endorsement rates, with tested inferences receiving significantly higher endorsement for control items than for target items (sound effects:  $\chi^2(1) = 13, p < .001$ ; emoji:  $\chi^2(1) = 5.4, p < .05$ ). As discussed above, the high endorsement of control as well as target items is unsurprising in UNEMBEDDED environments, since there is no logical operator for the inference to interact with in the first place. As for the



**Figure 4** Results of inferential judgment task with sound effects; dots represent individual participants' mean ratings.



**Figure 5** Results of inferential judgment task with emoji; dots represent individual participants' mean ratings.

difference between targets and controls, Tieu et al. speculate that the higher endorsement of inferences for control items stems from a difference between targets and controls in terms of whether the gesture (or, in our case, sound effect or emoji) can be ignored. In the target items, the sentence is perfectly interpretable if the co-linguistic content is ignored entirely. This stands in contrast to the control items, where deictic “this” explicitly refers to the interpretation of the gesture, sound effect, or emoji. Since the gesture/sound effect/emoji is responsible for the inference being tested, ignoring it in the target condition would lead to lower inference endorsement rates.

In the case of MIGHT, in all three studies both target and control items had relatively high endorsement rates—a somewhat surprising result for the controls—with target items receiving significantly higher endorsement rates for gestures and emoji (emoji:  $\chi^2(1) = 12, p < .001$ ), but not for sound effects ( $\chi^2(1) = .42, p = .52$ ). The difference between targets and

controls for gestures and emoji suggests that co-linguistic gesture- and emoji-derived inferences interact with “might” in a manner different from at-issue modifiers like “[in this direction]<sub>UP</sub>”. We are as yet unsure what led to the distinction between sound effects on the one hand and gestures and emoji on the other, though we suspect it is tied to the more general phenomenon of surprisingly high endorsement rates for MIGHT control items, which might have eliminated the difference between controls and targets.

For NEGATION, the tested inferences were endorsed more strongly for target sentences than for controls across all three studies (sound effects:  $\chi^2(1) = 20, p < .001$ ; emoji:  $\chi^2(1) = 37, p < .001$ ). This suggests that co-linguistic gestures, sound effects, and emoji all give rise to conditional inferences in negated environments, again unlike at-issue modification.

**EACH** As discussed previously, in the EACH environment both existential and universal inferences were predicted to receive high ratings for both control and target items. This was indeed the case across all three studies.

**NONE** In the NONE environment, target items were predicted to receive significantly higher endorsements than control items for both existential and universal inferences. The results for NONE are remarkably similar across the three experiments, and provide strong support for the presence of universal conditional inferences with co-linguistic content, distinguishing it from at-issue modification. In all three experiments, the strength of the effect was the same for both existential and universal inferences, with no significant interaction observed between Condition (target vs. control) and Reading (existential vs. universal) (sound effects:  $\chi^2(1) = .0078, p = .93$ ; emoji:  $\chi^2(1) = .43, p = .51$ ). Following Tieu et al.’s reasoning, this suggests that the existential inference is likely a consequence of the stronger universal inference; if the existential inference were derived independently, it should have strictly speaking been more strongly endorsed than the universal inference. The results for NONE therefore provide evidence for *universal conditional inferences* across the board.

**EXACTLY ONE** As discussed above, for EXACTLY ONE, control items were expected to receive high existential and low universal endorsement. Thus, with respect to existential inferences we observe similar results as in the UNEMBEDDED and EACH environments: inferences were strongly endorsed for both target and control items, and were actually more strongly endorsed for the controls in the gesture and sound effect experiments (sound effects:  $\chi^2(1) = 4.7, p < .05$ ), an observation that can again be attributed to differences in the feasibility of ignoring co-linguistic content, as in the UNEMBEDDED cases.

Meanwhile, since the universal inference is expected to receive low endorsement in control items, significantly higher endorsement in the gesture, sound effect, and emoji targets compared to their associated controls presents strong evidence that all three generate universal inferences under “exactly one” that distinguish them from at-issue modification (sound effects:  $\chi^2(1) = 9.5, p < .01$ ; emoji:  $\chi^2(1) = 29, p < .001$ ).



### 3.4 Summary

To summarize, the results from our sound effect and emoji experiments, in conjunction with the gestural results from Tieu et al. (2018), suggest that inferences from co-speech sound effects and co-text emoji interact with spoken logical operators in the same manner as gestures, in ways that can often be clearly differentiated from at-issue modifiers like “[in this direction]<sub>UP</sub>”. This includes conditional inferences arising in negation and modal environments, as well as universal conditional inferences in quantificational environments. In certain cases where both control and target sentences were predicted to have high endorsement rates, the control sentences received slightly higher endorsements; we follow Tieu et al. (2018) in attributing this to the observation that the target examples could easily be interpreted while ignoring the co-linguistic content entirely, unlike the control examples in which deictic spoken material (namely, “this”) explicitly referred to the semantic interpretation of the co-linguistic material. The one substantive contrast between experiments was that whereas a significant difference favoring target items in the MIGHT environment was found for gestures and emoji, this difference was not found for sound effects. We have left a full account of this result open, but speculated that it was tied to something more general that arose in all three experiments: namely, a surprisingly high endorsement rate for control sentences in the MIGHT environment, something that itself ought to be addressed in future work on this topic.

## 4 Conclusion

In this paper we have reported experimental evidence that co-speech sound effects and co-text emoji interact with logical operators in the same manner that gestures do. This claim, if true, substantially constrains the hypothesis space as far as what kinds of secondary content behave in a gesture-like fashion, as well as the related question of what it is about co-speech gestures that makes them semantically behave in the way that they do. The results favor less restrictive hypotheses like **co-linguistic content** and **secondary content** over more restrictive hypotheses like **co-speech gesture uniqueness**, **multimodality**, and **embodied content**.

It is worth noting that Esipova (2019) has argued that the nature of the semantic contribution of co-speech gestures, including how the inferences derived therefrom interact with spoken logical operators, depends on the type of syntactic constituent that the gesture is interpreted as modifying: for example, she argues that gestures that modify noun phrases like “the girl” make different contributions from gestures modifying verb phrases like “use the stairs”. The studies reported in this paper focused exclusively on gestures, sound effects, and emoji that are aligned with (and presumably interpreted as modifying) the verb phrase. While it seems to us unlikely that sound effects and emoji should pattern so closely with gestures when it comes to verb phrase modification while patterning entirely differently with respect to noun phrase modification, the cross-categorical similarity in inferential behavior between gestures, sound effects, and emoji still warrants further experimental investigation. Furthermore, the studies reported by Tieu et al. (2018) and in this paper focused exclusively

on *iconic* co-linguistic content, i.e., content with a non-arbitrary mapping between form and meaning: for example, the emoji 🚬 has a meaning related to cigarettes because the emoji is itself a depiction of a cigarette. This can be contrasted with *emblematic* content, which has a more arbitrary relation between form and meaning, such as the “thumbs up” gesture or the 👍 emoji (often used to indicate strong approval or support). The similarities and differences between iconic and emblematic content inferences with respect to interactions with logical operators are yet to be fully established for either gesture or other content; this constitutes another domain in which future work ought to compare and contrast gestures and non-gestures. Finally, content of a single general kind can apparently vary in the types of semantic interpretation it conveys; see, e.g., Grosz et al.’s (to appear) work on *face emoji* like 😊 versus *activity emoji* like the bicycle emoji 🚲. There is substantial work that needs to be done in order to assess the full range of possible types of interpretations for gestures, sound effects, and emoji, as well as how variation in the types of interpretations might lead to variation in the ways in which these interpretations interact with logical operators. As research in this area proceeds, it will be important to keep an eye on points of possible convergence and divergence between various kinds of inference-inducing content.

It is also worth noting that even if it is definitively established that sound effects and emoji behave identically to gestures across the board, this observation does not conclude the task of determining what kinds of content behave in this manner. For example, the choice between the **co-linguistic content** and **secondary content** hypotheses is still unresolved: co-speech gestures and sound effects and co-text emoji all feature linguistic material (in the narrow sense) as the primary content, so whether the inference pattern also extends to secondary content where the primary content is not speech, text, or sign is as yet unresolved. One way of testing this would be to study cases where *pro-speech* gestures of the sort mentioned above have secondary, presumably non-manual gestures aligned with them, what one might call *co-gesture gestures*. One possible example from Esipova (2019), in which a manual pro-speech gesture comes with a seemingly secondary non-manual gesture, was discussed in fn. 6; however, see fn. 7 for discussion of challenges involved in investigating such cases. It is our hope that these challenges can eventually be surmounted, as co-gesture gestures could prove valuable in more precisely determining the class of content that behaves like co-speech gestures.

Finally, even if the class of secondary content that behaves in a gesture-like manner is eventually fully determined, this does not in and of itself tell us everything of relevance about the cognitive mechanisms that underlie this shared behavior. For example, if a less restrictive hypothesis turns out to be correct, one plausible explanation would be that gestures genuinely are not special as far as this aspect of human cognition is concerned: there is some cognitive mechanism that generally serves to integrate co-linguistic or secondary material in the appropriate fashion, regardless of what that material is (see, e.g., Pasternak 2021). But another, equally plausible explanation would be that gestures in fact *are* special: the cognitive mechanism at play is “designed” specifically for gestures, but other types of content can in some way be coerced into exploiting this mechanism in order to generate gesture-like inferences. Simply determining the range of content that gives rise to gesture-like inferences

cannot in and of itself favor one hypothesis over the other. Instead, finer-grained experimental methods are required to tease apart these hypotheses.

## Contributions

R.P. and L.T. share joint lead authorship of this work. R.P. and L.T. designed and implemented research; L.T. analyzed data; and R.P. and L.T. wrote the paper.

## Acknowledgments

We wish to thank Emmanuel Chemla, Cornelia Ebert, Francesco Pierini, Philippe Schlenker, and audience members at the ZAS workshop *Linguistic investigations beyond language*, the 2020 Australian Linguistic Society Annual Conference, and Linguistic Evidence 2020 for helpful discussion. Comments from Maria Esipova and an anonymous reviewer were extremely valuable.

## Funding

**The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article:** The research leading to this work was supported by Western Sydney University through the University's Research Theme Champion support funding. R.P.'s research is funded by DFG grant #387623969 (*DPBorder*, PIs: Artemis Alexiadou and Uli Sauerland).

## Declaration of conflicting interests

The Authors declare that there is no conflict of interest.

## References

- Anvari, Amir. 2017. Dislocated co-suppositions. In Alexandre Cremers, Thom van Gessel & Floris Roelofsen (eds.), *Proceedings of the 21st Amsterdam Colloquium*, 106–114. Amsterdam: ILLC.
- Aristodemo, Valentina. 2017. *Gradable constructions in Italian Sign Language*. Paris: École des Hautes Études en Sciences Sociales dissertation.
- Bates, D., M. Mächler, B. Bolker & S. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48.
- Byrne, R. W., E. Cartmill, E. Genty, K. E. Graham, C. Hobaiter & J. Tanner. 2017. Great ape gestures: intentional communication with a rich set of innate signals. *Animal Cognition* 20(4). 755–769.
- Chemla, Emmanuel. 2009. Presuppositions of quantified sentences: experimental data. *Natural Language Semantics* 17(4). 299–340.

- Chemla, Emmanuel & Lewis Bott. 2013. Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes* 28(3). 241–260.
- Chemla, Emmanuel & Philippe Schlenker. 2012. Incremental vs. symmetric accounts of presupposition projection: an experimental approach. *Natural Language Semantics* 20(2). 177–226.
- Davidson, Kathryn. 2015. Quotation, demonstration, and iconicity. *Linguistics and Philosophy* 38(6). 477–520.
- Domaneschi, Filippo, Elena Carrea, Carlo Penco & Alberto Greco. 2013. The cognitive load of presupposition triggers: mandatory and optional repairs in presupposition failure. *Language and Cognitive Processes* 29(1). 136–146.
- Ebert, Cornelia & Christian Ebert. 2014. Gestures, demonstratives, and the attributive/referential distinction. Slides from a talk given at Semantics and Philosophy in Europe (SPE 7).
- Ebert, Cornelia, Stefan Evert & Katharina Wilmes. 2011. Focus marking via gestures. In Ingo Reich, Eva Horch & Dennis Pauly (eds.), *Proceedings of Sinn und Bedeutung 15*, 193–208. Saarbrücken, Germany: Saarland University Press.
- Emmorey, Karen. 1999. Do signers gesture? In Lynn S. Messing & Ruth Campbell (eds.), *Gesture, speech, and sign*, 133–159. Oxford: Oxford University Press.
- Esipova, Maria. 2018a. Composition and projection of adnominal content across modalities. New York University, Ms.
- Esipova, Maria. 2018b. Focus on what's not at issue: Gestures, presuppositions, appositives under contrastive focus. In Uli Sauerland & Stephanie Solt (eds.), *Proceedings of Sinn und Bedeutung 22*, 385–402. Berlin: ZAS.
- Esipova, Maria. 2019. *Composition and projection in speech and gesture*. New York, NY: New York University dissertation.
- Gawne, Lauren & Gretchen McCulloch. 2019. Emoji as digital gestures. *Language@Internet* 17. Article 2.
- Geurts, Bart. 1998. Presuppositions and anaphors in attitude contexts. *Linguistics and Philosophy* 21(6). 545–601.
- Gillespie, Maureen, Ariel N. James, Kara D. Federmeier & Duane G. Watson. 2014. Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition* 132(2). 174–180.
- Goldin-Meadow, Susan & Diane Brentari. 2017. Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences* 40. e46.
- Goldin-Meadow, Susan, Howard Nusbaum, Spencer D. Kelly & Susan Wagner. 2001. Explaining math: Gesturing lightens the load. *Psychological Science* 12(6). 516–522.
- Grosz, Patrick Georg, Elsi Kaiser & Francesco Pierini. to appear. Discourse anaphoricity and first-person indexicality in emoji resolution. To appear in *Proceedings of Sinn und Bedeutung 25*.
- Heim, Irene. 1983. On the projection problem for presuppositions. In *Proceedings of the second west coast conference on formal linguistics*, 114–125. Stanford, CA: Stanford University Press.
- Heim, Irene. 1992. Presupposition projection and the semantics of attitude verbs. *Journal of Semantics* 9(3). 183–221.
- Hunter, Julie. 2019. Relating gesture to speech: reflections on the role of conditional presuppositions. *Linguistics and Philosophy* 42(4). 317–332.
- Iverson, Jana M. & Susan Goldin-Meadow. 1997. What's communication got to do with it? gesture in children blind from birth. *Developmental Psychology* 33(3). 453–467.
- Iverson, Jana M. & Susan Goldin-Meadow. 1998. Why people gesture when they speak. *Nature* 396.

- 228.
- Iverson, Jana M. & Susan Goldin-Meadow. 2005. Gesture paves the way for language development. *Psychological Science* 16(5). 367–371.
- Karttunen, Lauri. 1973. Presuppositions of compound sentences. *Linguistic Inquiry* 4(2). 169–193.
- Krauss, Robert M. 1998. Why do we gesture when we speak? *Current Directions in Psychological Science* 7(2). 54–60.
- Lascarides, Alex & Matthew Stone. 2009. A formal semantic analysis of gesture. *Journal of Semantics* 26(3). 393–449.
- Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(1). 339–359.
- Pasternak, Robert. 2019. The projection of co-speech sound effects. Leibniz-Center for General Linguistics (ZAS), Ms. <https://ling.auf.net/lingbuzz/004520>.
- Pasternak, Robert. 2021. Co- and pro-speech integration: The parsing hypothesis. Leibniz-Center for General Linguistics (ZAS), Ms. <https://ling.auf.net/lingbuzz/005906>.
- Pierini, Francesco. to appear. Emojis and gestures: a new typology. To appear in *Proceedings of Sinn und Bedeutung* 25.
- Potts, Christopher. 2005. *The logic of conventional implicatures*. Oxford: Oxford University Press.
- R Core Team. 2016. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Schlenker, Philippe. 2009. Local contexts. *Semantics & Pragmatics* 2(3). 1–78.
- Schlenker, Philippe. 2018a. Gesture projection and cosuppositions. *Linguistics and Philosophy* 41(3). 295–365.
- Schlenker, Philippe. 2018b. Iconic pragmatics. *Natural Language & Linguistic Theory* 36(3). 877–936.
- Schwarz, Florian. 2007. Processing presupposed content. *Journal of Semantics* 24(4). 373–416.
- Schwarz, Florian. 2015. *Experimental perspectives on presuppositions (studies in theoretical psycholinguistics 45)*. Cham: Springer International Publishing.
- Schwarz, Florian. 2019. Presuppositions, projection, and accommodation - theoretical issues and experimental approaches. In Chris Cummins & Napoleon Katsos (eds.), *Handbook of experimental semantics and pragmatics*, 83–113. Oxford: Oxford University Press.
- Schwarz, Florian & Sonja Tiemann. 2017. Presupposition projection in online processing. *Journal of Semantics* 34(1). 61–106.
- Stalnaker, Robert. 1973. Presuppositions. *Journal of Philosophical Logic* 2(4). 447–457.
- Strawson, P. F. 1950. On referring. *Mind* 59(235). 320–344.
- Tieu, Lyn, Robert Pasternak, Philippe Schlenker & Emmanuel Chemla. 2017. Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa: a journal of general linguistics* 2(1): 102. 1–27.
- Tieu, Lyn, Robert Pasternak, Philippe Schlenker & Emmanuel Chemla. 2018. Co-speech gesture projection: Evidence from inferential judgments. *Glossa: a journal of general linguistics* 3(1): 109. 1–21.
- van der Sandt, Rob. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9(4). 333–377.
- Wilmes, Katharina. 2009. *Hands in Focus: Focus Marking by Speech Accompanying Gestures*. Bachelor's thesis, University of Osnabrück.
- Zlogar, Christina & Kathryn Davidson. 2018. Effects of linguistic context on the acceptability of co-speech gestures. *Glossa* 3(73). 1–28.