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

Robert Pasternak[a,1,2] **and Lyn Tieu**[b,c,d,e,1]

[a]Leibniz-Center for General Linguistics (ZAS), Schützenstraße 18, 10117 Berlin, Germany; [b]Office of the Pro Vice-Chancellor (Research and Innovation), Western Sydney University, Penrith NSW 2751, Australia; [c]School of Education, Western Sydney University, Penrith NSW 2751, Australia; [d]MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Penrith NSW 2751, Australia

This manuscript was compiled on March 12, 2020

Co-speech gestures have been argued to trigger inferences that "project" in certain linguistic environments, i.e., interact with logical operators in particular ways. Despite recent theoretical and experimental interest in the question of *how* these inferences project, a little-addressed question is *why* gestural inferences project in the way that they do. We present two experiments investigating sentences with *co-speech sound effects* and *co-text emoji* in lieu of gestures, revealing a remarkably similar projection pattern to that of co-speech gestures. The results suggest that gestural inferences do not project because of any traits specific to gestures, but rather because of their status as *co-linguistic content*.

co-linguistic content | gesture | emoji | semantics | pragmatics

Co-speech gestures—gestures temporally aligned with spoken content—serve a variety of purposes, including lightening the cognitive burden of language production (1–3); marking information-structural features like focus (4, 5); and contributing additional meaningful content to the speaker's utterance. An instance of the latter use of gestures can be seen in example 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]$_{\text{UP}}$.
⇝ The girl will go up the stairs.

Recent years have seen growing interest in this last use of gestures, and in the precise nature of the meaningful contributions these gestures make (6–14). As will be discussed in the next section, much of this research has focused on the so-called *projection* of gestural inferences, i.e., how they interact with logical operators in the spoken sentence. But while plenty of work has explored *how* gestural inferences project, considerably less has asked *why* they project as they do. 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.

In this paper we will restrict this hypothesis space 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. Crucially, both co-speech sound effects and co-text emoji lack many of the distinguishing traits of their gestural counterparts, meaning that if the same inference patterns are found for sound effects and emoji as for gestures, each of these gesture-specific traits can be excluded as a possible source for the associated inference pattern. In fact, we will report experimental evidence from inferential judgment tasks suggesting precisely this, as inferences from co-speech sound effects and co-text emoji are shown to behave in a fashion strikingly similar to that observed in a previous study using the same task to explore gestural inferences (10). As a result we conclude that the apparent "gestural" inference pattern most likely encompasses a much broader array of content, and thus might be more accurately referred to as the *co-linguistic* inference pattern.

Background

A growing body of work investigates how inferences from co-speech gestures *project*. Before discussing the projection of gestural inferences in particular, a more general overview of projection is in order.

Projection. Among other things, the gesture-less example 2a leads one to infer (I) that Eve does not currently smoke, and (II) that Eve previously smoked. When sentential negation (*not*) is added, as in example 2b, the "does not currently smoke" inference disappears, an unsurprising result of the meaningful contribution of negation. But we still infer that Eve previously smoked: this inference *projects through negation*, unlike the "does not currently smoke" inference.

(2) a. Eve stopped smoking.
b. Eve did not stop smoking.

The inference in examples 2a and 2b that Eve previously smoked is a *presupposition*, so called because it seems that it must be true in order

Significance Statement

Gestures often contribute meaningful content to the utterances with which they coincide. Interestingly, inferences from such co-speech gestures have been found to interact with logical operators in the spoken portion of the sentence, in ways previously assumed to be confined to language (*projection*). While gestures exhibit several unique traits that could conceivably give rise to this behavior, we report experimental evidence that the same pattern arises with co-speech sound effects and co-text emoji, which lack many of the traits that make gestures special. These findings suggest that gestures are just one instance of a much broader phenomenon of *co-linguistic content*, in which a diverse array of non-linguistic inferences can integrate into and interact with linguistic meaning.

Author contributions: R.P. and L.T. designed and implemented research; L.T. analyzed data; and R.P. and L.T. wrote the paper.

The authors have no conflict of interest.

[1]R.P. and L.T. contributed equally to this work and share joint first authorship.

[2]To whom correspondence should be addressed. E-mail: mail@robertpasternak.com

*Note that we are concerned only with gestures coinciding with auditory speech, thereby putting aside sign languages. Sign languages are distinct from gestures along multiple dimensions, and thus should not be grouped with them (15). We also set aside the possibility of "co-sign" gestures, though cases of gesture and speech occupying the same (visual) modality could be of significant interest.

for an utterance of the sentence to be felicitous: if Eve never smoked, saying 2a or 2b is often odd (16). The projection through negation seen in example 2b is thus *presupposition projection*.

Projection is not merely a yes-or-no matter, as can be seen by exploring interactions with quantifiers like *none*. When the subject in 2a (*Eve*) is replaced with the quantificational *none of these three students*, as in example 3, the presupposition of having previously smoked projects *universally*: we derive an inference that *each* student previously smoked (17).

(3) None of these three students stopped smoking.

It is worth noting that presuppositions do not always project, even in environments from which they often do: they can instead be *locally accommodated*, in which case the presupposed content essentially loses its presuppositional character and simply becomes part of the sentence's assertion (18). Example 4 illustrates this:

(4) None of these three students stopped smoking, since none of them smoked to begin with.

If the presupposition in the first clause projected universally like in example 3, we would arrive at a contradiction: each student used to smoke (first clause), and none did (second clause). Instead, the interpretation of example 4 seems similar to that of example 5:

(5) None of these three students both previously smoked and stopped smoking, since none of them smoked to begin with.

In other words, the "previously smoked" inference is incorporated into the assertive component (the "regular meaning") of the sentence, rather than being presupposed.

Not all presuppositions project identically, and presuppositions are not the only type of linguistic inference that projects, with myriad grammatical and other factors influencing how inferences project in various environments. These subtleties do not concern us for our present purposes. Instead, we simply note three main takeaways: (I) some inferences project, and some do not; (II) projection is not merely yes-or-no, as evidenced by the interaction between projecting content and quantifiers like *none*; and (III) sometimes content that otherwise could project does not.

Schlenker on Gesture Projection. Recent work has explored if and how gesture-derived inferences project, typically operating under the assumption—by no means *a priori* obvious—that gesture projection follows some pattern also seen in spoken language. Since our concern in this paper is not how to characterize the projection pattern but rather what kinds of content give rise to it, for us the important issue is not the nature of the theoretical proposals that have been put forward, but the available empirical evidence on gesture projection. Here we go over the intuitive judgments offered by Schlenker (*ref.* 8), as those are the ones tested by Tieu *et al.* (*refs.* 9, 10); later we will discuss Tieu *et al.*'s inferential judgment task (IJT) study confirming that Schlenker's gesture judgments are shared by untrained speakers (10), in addition to our own IJTs for sound effects and emoji.

Schlenker proposes that when a gesture G coincides with some spoken content V, what results is a conditional inference that he refers to as a **cosupposition**: roughly, that "V-ing entails G-ing". For instance, *[use the stairs]$_{UP}$* generates an inference that using the stairs entails going upward. In unembedded environments like example 1, this leads to a plain conjunctive inference: the spoken utterance asserts that the girl will use the stairs, and the gesture generates an inference that using the stairs entails going upward, so by combining these two inferences we get that the girl will take the stairs upward.

Interestingly, this cosuppositional inference seems to project through a variety of logical operators. Take, for instance, sentential negation, as shown in example 6:

(6) The girl will not [use the stairs]$_{UP}$.
⇝ If the girl were to use the stairs, she would go up the stairs.

If the cosuppositional inference did not project through negation, we would expect example 6 to simply mean that the girl will not go up the stairs, as in the seemingly similar example 7, where *this* explicitly refers to the content of the gesture:

(7) The girl will not use the stairs [in this direction]$_{UP}$.

But notice that example 7 is true if the girl will go down the stairs, while 6 says something stronger: the girl will not use the stairs *at all*, and if she did, she would go up. Thus, the inference that using the stairs entails going upward "survives": it projects through negation.

Similar observations can be made with the modal verb *might*:

(8) The girl might [use the stairs]$_{UP}$.
⇝ If the girl were to use the stairs, she would go up the stairs.

Without projection through *might*, example 8 should mean only that the girl might go up the stairs, much like in example 9:

(9) The girl might use the stairs [in this direction]$_{UP}$.

But 8 again seemingly makes a stronger claim than 9: if the girl also might go *down* the stairs, then only the weaker 9 is true. Once again, this is due to projection: example 8 leads one to infer that the girl might take the stairs, and that if she does, she will go up. That is, she will either take the stairs upward or not take them at all. The conditional inference thus projects through *might* as well.

Schlenker also discusses cases with quantificational subjects, such as *each*, *none*, and *exactly one*. Recall from the prior discussion that when quantifiers are introduced, projection is no longer a mere "yes or no" matter like it is with *might* or negation. In the specific case of gestures, Schlenker argues that the cosuppositional inference projects universally, analogous to the projecting inference in example 3. This is indicated in example 10:

(10) a. Each of these three girls will [use the stairs]$_{UP}$.
b. None of these three girls will [use the stairs]$_{UP}$.
c. Exactly one of these three girls will [use the stairs]$_{UP}$.
⇝ For each of these three girls, if she were to use the stairs, she would go up the stairs.

In summary, Schlenker (*ref.* 8) argues that a gesture G coinciding with spoken content V generates a cosuppositional inference that "V-ing entails G-ing", and that this inference projects through *might* and negation, and projects universally in quantificational environments. Tieu *et al.* (*ref.* 10) provide experimental evidence that untrained speakers share Schlenker's judgments; we will return to their experiment when we discuss our own IJT experiments with sound effects and emoji.

Possible Sources of the Inference Pattern. Temporarily assuming that Schlenker's characterization of co-speech gesture projection is accurate, an important and thus far unanswered followup question is what it is about co-speech gestures that gives rise to this particular projection profile. Though there is obviously a large space of conceivable possibilities, in a short theoretical paper Pasternak (*ref.* 19) offers what seem to be the four most plausible hypotheses, which we refer to as **gesture uniqueness**, **multimodality**, **embodied co-speech content**, and **general co-linguistic content**.

Gesture Uniqueness. 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 (20), and are so deeply embedded in the human language apparatus 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 (21, 22). Furthermore, non-human primates have been shown to make rich use of communicative gestures (23), suggesting that gesture may have played an important evolutionary role in interhuman communication. Given their linguistic importance, it may be that this projection profile is specific to gestures and is somehow tied to their unique status in human language evolution, acquisition, or day-to-day use.

Multimodality. 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-gesture pairings. If this is the case, we might expect the relevant projection profile to be exhibited not only by gestures, but also by other types of visual co-speech content like pictures, animations, or signage.

Embodied Co-Speech Content. A third possibility is that the co-speech projection pattern is neither specific to gestures nor a result of multimodal information transferral, but is instead a feature of what might be called *embodied co-speech content*: co-speech content that is in some sense directly produced by the speaker's body. Unlike the multimodality hypothesis, this would exclude visual co-speech information not produced by the speaker's body (e.g., animations), but it would include embodied content that occupies the same (auditory) modality as speech, such as vocal modulations and clapping noises.

General Co-Linguistic Content. This hypothesis is the least restrictive: co-speech gestures project as they do not because of any traits that are specific to gestures, but simply because they are co-linguistic content. According to this hypothesis, all else being equal we expect the gesture projection pattern to arise for a wide variety of co-linguistic content, including those kinds mentioned above, as well as sound effects and emoji, to which we now turn.

What Sound Effects and Emoji Could Tell Us. There is an obvious path to teasing apart these and other potential hypotheses: try to replicate the results from previous gesture projection studies, but replace gestures with other types of co-linguistic content that are compatible with only a proper part of the hypothesis space. For this reason, Pasternak (*ref.* 19) suggests using *co-speech sound effects*, communicative sound effects aligned with spoken content. This is illustrated in Pasternak's example 11, 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.)

(11) The soldier will [assassinate his target]$_{\texttt{explode}}$.

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. In other words, none of the first three hypotheses above predicts that inferences from sound effects should project like those from gestures. Nonetheless, Pasternak argues that co-speech sound effects do in fact project like gestures, citing as evidence speech-sound pairings like the one in example 12:

(12) The soldier will not [assassinate his target]$_{\texttt{explode}}$.
⇝ If the soldier were to assassinate his target, he would do so via explosion.

Another potential test for these four hypotheses is *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 (24). With respect to projection, Pierini (*ref.* 25) 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 example 13, in our examples both in this paper and in the reported experiment, emoji "bracket" the verb phrase on either side, as in example 14:

(13) The student will step out of the classroom 🚽

(14) 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 (*ref.* 8) discusses, *post*-speech gestures—gestures following and not coinciding with speech—project 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.

Example 14 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 projection pattern seems to arise with emoji too. For instance, example 15 with sentential negation gives rise to a counterfactual inference parallel to the gestural inference in 6 and the sound effect inference in 12:

(15) 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 project like gestures. If Pierini's observations are correct, then emoji further point toward a more general co-linguistic projection pattern.

In arguing that sound effects and emoji project like gestures, Pasternak and Pierini follow Schlenker in relying on intuitive judgments, a common methodology for research in theoretical linguistics. However, judgments about inferences from gestures, sound effects, and emoji are often subtle and unstable, so these claims must be more rigorously tested before deciding on the shape of the broader empirical picture. With this in mind, in the next section we discuss three experimental studies performed using an inferential judgment task: the gesture experiment reported by Tieu *et al.* (*ref.* 10), and our own studies on sound effects and emoji.

Experiments

Tieu *et al.*'s Gesture Study. Tieu *et al.* (*ref.* 10) use an inferential judgment task (IJT) to test the projection of gestural inferences in six environments: UNEMBEDDED (cf. 1), MIGHT (cf. 8), NEGATION (cf. 6), 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 along with a co-speech gesture; text was provided below the video indicating a particular potential inference, and participants were asked to indicate with a sliding 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, 8, and 6, 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**, illustrated in example 16:

(16) a. **Existential:** For at least one of these three girls, if she were to use the stairs, she 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]*$_{UP}$ was replaced with *use the stairs [in this direction]*$_{UP}$. Recall that in such examples projection of the directional inference is neutralized, hence the differences in interpretation between examples 6 and 7, and between 8 and 9. Target vs. control was a between-subject factor: one group of participants saw only target items, and one group only control items.

It is clear from the preceding discussion that in the MIGHT and NEGATION environments, participants were expected to give higher ratings in the target condition, compared to relatively low ratings in the control condition. But note that low endorsement for control items is not to be expected across all environments. Take, for example, the control sentence in the UNEMBEDDED environment:

(17) The girl will use the stairs [in this direction]$_{\text{UP}}$.

Because the UNEMBEDDED environment has no intervening logical operator—and thus nothing to differentiate between projection and non-projection—this sentence should in fact generate the inference that the girl will go up the stairs, just like target example 1. We would thus predict high inference endorsement rates for both control and target sentences. Now consider the control sentence for EACH:

(18) Each of these three girls will use the stairs [in this direction]$_{\text{UP}}$.

The semantic interpretation of *each* inherently involves universal quantification. Thus, even when the projection of the gestural inference is stymied in control items, the universal quantification in the semantics of *each* nonetheless generates a universal (and thus also existential) inference. We would therefore expect high endorsement rates for both the existential and universal inferences in the EACH control condition. Finally, consider EXACTLY ONE:

(19) Exactly one of these three girls will use the stairs [in this direction]$_{\text{UP}}$.

Intuitively, the semantics of *exactly one* is equivalent to that of *at least one and no more than one*. It thus entails existential quantification (*at least one*), but contradicts universality (*no more than one*). As a result, for control items in the *exactly one* environment, we would expect high endorsement of existential inferences and low endorsement of universal inferences.

In summary, for control items endorsement is expected to be high in the UNEMBEDDED environment, for both existential and universal inferences with EACH, and for existential inferences with EXACTLY ONE, and otherwise relatively low. For target items, meanwhile, the rate of inference endorsement is predicted to depend on the co-speech gesture's projection profile, with the most informative cases being those in which endorsements for control items are low, as a significant difference between target and control items would constitute clear evidence of projection.

The results of Tieu *et al.*'s study can be seen in Fig. 1, where red bars ("Gesture") indicate mean endorsement for target items, and blue bars ("Asserted") indicate mean endorsement for controls. We will discuss further details of Tieu *et al.*'s results alongside those of our own studies on co-speech sound effects and emoji.

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.*: on each trial, participants either listened to an embedded SoundCloud audio file of a sentence containing a sound effect (Experiment 1) or read a sentence containing emoji (Experiment 2), and then had to use a slider to indicate the degree to which they drew the inference indicated in text below the test item. 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 (example 20) and emoji (example 21) equivalents of the gestural targets and controls of Tieu *et al.* (*ref.* 10). Tested inferences were also constructed in a parallel fashion (example 22).

(20) a. **Target:** The student will not [step out of the classroom]$_{\texttt{flush}}$.

b. **Control:** The student will not step out of the classroom to do this: []$_{\texttt{flush}}$.

(21) 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: 🚽

(22) 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. In Experiment 1, 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; in Experiment 2, we tested five distinct emoji: cigarette (🚬), rain-on-umbrella (☔), phone (📞), plane (🛫), and toilet (🚽). To ensure that the 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.

In each experiment, 200 MTurk workers were randomly assigned to either the target or control condition, for a total of 100 participants per condition. 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.

Results: Sound Effects and Emoji Project Like Gestures. The raw data and R scripts for analysis (including details of statistical models) can be accessed at https://osf.io/5vh7m. Figures 2 and 3 show very similar endorsement rates to those for gestures in Tieu *et al.*'s Figure 1. Impressionistically, across the three types of 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 presence of the inference was specifically due to projection. We thus used mixed effect linear regression models to determine if Condition (TARGET VS. CONTROL) was a significant predictor of the inferential judgment responses.

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 due to the lack of a logical operator to distinguish between projection and non-projection. As for the 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, in contrast to the control items, where *this* explicitly refers to the interpretation of the co-linguistic content.

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 projection of the co-linguistic content's inference through *might*. 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.

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 gesture-, sound effect-, and emoji-derived conditional inferences all project through negation.

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, since the semantics of *each* entails universal quantification. This was indeed the case across all three studies.

NONE. In the NONE environment, both existential and universal inferences were expected to receive low endorsement for control items, meaning that a significant difference between target and control endorsement rates would be indicative of projection. The results for NONE are remarkably similar across the three experiments, and provide evidence that both universal and existential inferences are contributed. 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* projection across the board.

EXACTLY ONE. For EXACTLY ONE, control items were expected to receive high existential and low universal endorsement, since the semantics of *exactly one* entails existential quantification (*at least one*) and contradicts universal quantification (*no more than one*). 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.

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 trigger universal inferences under *exactly one* (sound effects: $\chi^2(1) = 9.5, p < .01$; emoji: $\chi^2(1) = 29, p < .001$).

Conclusion

Overall, our results for both sound effects and emoji pattern quite closely with those reported by Tieu *et al.* for gestures: evidence was found for projection through negation and *might*, as well as universal projection from quantificational environments. The extension of the observed gesture projection pattern to sound effects and emoji indicates that this projection profile is not specific to gestures, nor does it arise from multimodality or embodied co-linguistic content. In other words, the evidence comes down strongly in favor of the **general co-linguistic content** hypothesis: the "co-speech gesture" projection profile is really a reflection of a much more general pattern of *co-linguistic content* projection.

ACKNOWLEDGMENTS. The research leading to this work was supported by Western Sydney University through the University's Research Theme Champion support funding. We wish to thank Emmanuel Chemla, Cornelia Ebert, Francesco Pierini, Philippe Schlenker, and audience members at

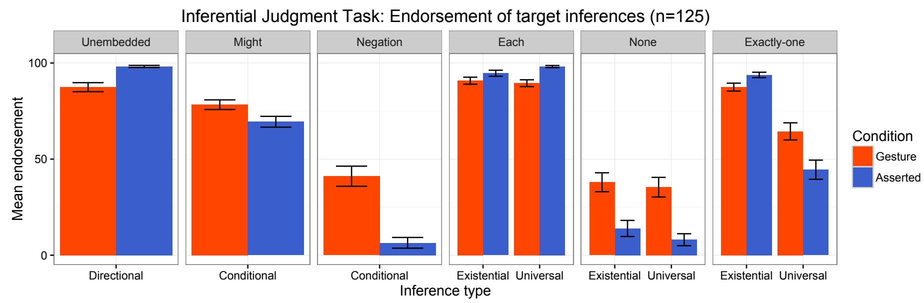

Fig. 1. Results of Tieu *et al.*'s (*ref.* 10) inferential judgment task with gestures.

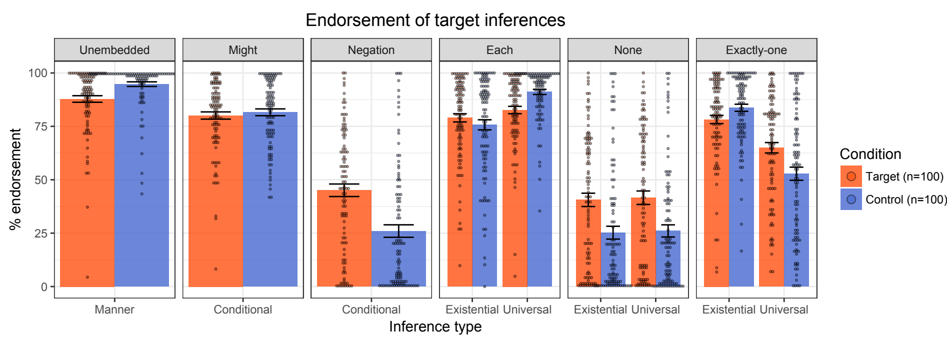

Fig. 2. Results of inferential judgment task with sound effects; dots represent individual participants' mean ratings.

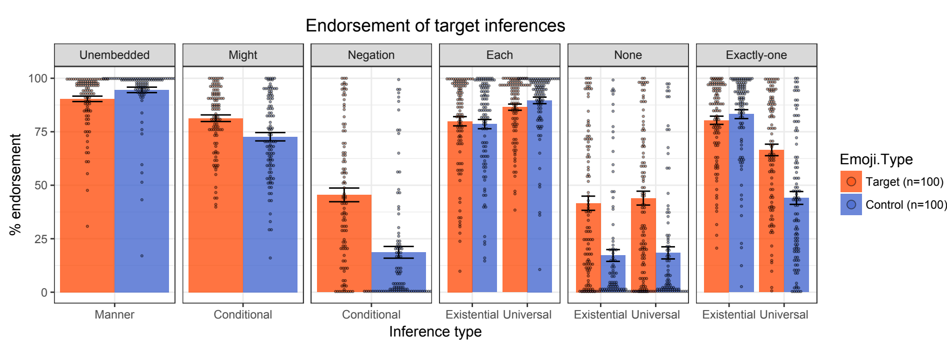

Fig. 3. Results of inferential judgment task with emoji; dots represent individual participants' mean ratings.

the ZAS workshop *Linguistic investigations beyond language* for helpful discussion. R.P.'s research is funded by DFG grant #387623969 (*DPBorder*, PIs: Artemis Alexiadou and Uli Sauerland).

1. Krauss RM (1998) Why do we gesture when we speak? *Current Directions in Psychological Science* 7(2):54–60.
2. Goldin-Meadow S, Nusbaum H, Kelly SD, Wagner S (2001) Explaining math: Gesturing lightens the load. *Psychological Science* 12(6):516–522.
3. Gillespie M, James AN, Federmeier KD, Watson DG (2014) Verbal working memory predicts co-speech gesture: Evidence from individual differences. *Cognition* 132(2):174–180.
4. Wilmes K (2009) *Hands in Focus: Focus Marking by Speech Accompanying Gestures*. Bachelor's thesis, University of Osnabrück.
5. Ebert C, Evert S, Wilmes K (2011) Focus marking via gestures in *Proceedings of Sinn und Bedeutung 15*, eds. Reich I, Horch E, Pauly D. (Saarland University Press, Saarbrücken, Germany), pp. 193–208.
6. Lascarides A, Stone M (2009) A formal semantic analysis of gesture. *Journal of Semantics* 26(3):393–449.
7. Ebert C, Ebert C (2014) Gestures, demonstratives, and the attributive/referential distinction. Slides from a talk given at Semantics and Philosophy in Europe (SPE 7).
8. Schlenker P (2018) Gesture projection and cosuppositions. *Linguistics and Philosophy* 41(3):295–365.
9. Tieu L, Pasternak R, Schlenker P, Chemla E (2017) Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa: a journal of general linguistics* 2(102):1–27.
10. 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):1–21.
11. Anvari A (2017) Dislocated co-suppositions in *Proceedings of the 21st Amsterdam Colloquium*, eds. Cremers A, van Gessel T, Roelofsen F. (ILLC, Amsterdam), pp. 106–114.
12. Esipova M (2018) Composition and projection of adnominal content across modalities. New York University, Ms.
13. Esipova M (2018) Focus on what's not at issue: Gestures, presuppositions, appositives under contrastive focus in *Proceedings of Sinn und Bedeutung 22*, eds. Sauerland U, Solt S. (ZAS, Berlin), pp. 385–402.
14. Hunter J (2018) Relating gesture to speech: reflections on the role of conditional presuppositions. *Linguistics and Philosophy* (Online First).
15. McNeill D (2005) *Gesture and Thought*. (University of Chicago Press, Chicago).
16. Strawson PF (1950) On referring. *Mind* 59(235):320–344.
17. Chemla E (2009) Presuppositions of quantified sentences: experimental data. *Natural Language Semantics* 17(4):299–340.
18. Heim I (1983) On the projection problem for presuppositions in *Proceedings of the Second West Coast Conference on Formal Linguistics*. (Stanford University Press, Stanford, CA), pp. 114–125.
19. Pasternak R (2019) The projection of co-speech sound effects. Leibniz-Center for General Linguistics (ZAS), Ms.
20. Iverson JM, Goldin-Meadow S (2005) Gesture paves the way for language development. *Psychological Science* 16(5):367–371.
21. Iverson JM, Goldin-Meadow S (1997) What's communication got to do with it? gesture in children blind from birth. *Developmental Pychology* 33(3):453–467.
22. Iverson JM, Goldin-Meadow S (1998) Why people gesture when they speak. *Nature* 396:228.
23. Byrne RW, et al. (2017) Great ape gestures: intentional communication with a rich set of innate signals. *Animal Cognition* 20(4):755–769.
24. Gawne L, McCulloch G (2019) Emoji as digital gestures. *Language@Internet* 17:Article 2.
25. Pierini F (2019) Emoji and gestures. Ecole Normale Supérieure, Ms.