RESEARCH

Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks

Lyn Tieu[1,2], Robert Pasternak[3], Philippe Schlenker[4] and Emmanuel Chemla[2]

[1] CCD, Department of Cognitive Science, Macquarie University, 16 University Avenue, Macquarie University 2109 NSW, AU

[2] Laboratoire de Sciences Cognitives et Psycholinguistique, Département d'études cognitives, ENS, EHESS, CNRS, PSL Research University, 29 rue d'Ulm, 75005 Paris, FR

[3] Stony Brook University, Stony Brook, NY 11794-4376, US

[4] Institut Jean-Nicod, Département d'études cognitives, ENS, EHESS, CNRS, PSL Research University, 29 rue d'Ulm, 75005 Paris, FR

Corresponding author: Lyn Tieu (lyn.tieu@gmail.com)

Two main analyses have been proposed to explain how co-speech gestures interact with logical operators. According to the Supplemental analysis (Ebert & Ebert 2014), co-speech gestures have the same semantic status as appositive relative clauses. According to the Cosuppositional analysis (Schlenker To appear a; b), co-speech gestures trigger a particular kind of presupposition. The sentence "John will not use the stairs", produced with an UP gesture (finger pointed upwards) is argued to give rise to the conditional presupposition that *if John were to use the stairs, he would use the stairs in an upwards trajectory*. Both the Supplemental and Cosuppositional analyses predict that inferences triggered by co-speech gestures should *project* out of the scope of operators, but not quite in the same way. We present an experimental investigation of the projection properties of the inferences arising from the co-speech gestures UP and DOWN in six different linguistic environments (plain affirmative and negative sentences, modal sentences containing "might", and quantified sentences containing "each", "none", and "exactly one"). Applying a reading detection analysis (Cremers & Chemla 2017) to the responses of a Truth Value Judgment Task and a Picture Selection Task, we find evidence for *existential projection* of the gestural inferences in the scope of "each", "none", and "exactly one", and, to some degree, *local accommodation* of the inferences. These results can be derived by the Cosuppositional analysis, in combination with an analysis of presupposition projection such as Beaver (2001), which predicts existential projection out of quantified structures; on the other hand, both findings are difficult to reconcile with the Supplemental analysis. Our projection results bring gestural inferences and verbal presuppositions closer together, but a remaining puzzle is why in quantified structures we obtain existential rather than universal inferences (Chemla 2009), the latter being the more standard finding in the presuppositional literature (though see Tieu et al. 2016 for evidence of universal projection of the same gestural inferences).

1 Gestures: Theoretical and empirical background

1.1 Gesture projection

Speakers often gesture spontaneously when producing spoken utterances. *Co-speech gestures*, in particular, co-occur simultaneously with spoken language expressions. Such gestures appear to enrich the spoken language utterance by depicting some aspect of the denoted situation. For instance, the sentence in (1), with the gesture UP in Figure 1 produced simultaneously with the verb "helped", appears to convey that John helped his

son by lifting him upwards (Schlenker To appear b). Here and elsewhere in the paper, we will indicate the spoken words that align with the gesture by placing them in square brackets.

(1)   John [helped]_LIFT his son.

In this paper, we will focus our attention on such *iconic* co-speech gestures, and the projection problem that they introduce: how do such gestures interact with the logical structure of the sentences that they co-occur with? For example, consider the sentence in (2), taken from Schlenker (To appear b), where the co-speech gesture LARGE (Figure 2) is embedded under the quantified expression "exactly-one."

(2)   Exactly one philosopher found [a bottle he liked]_LARGE.

Figure 1: The co-speech gesture LIFT, accompanying the verb "helped" in (1). Arrows indicate the upwards motion of the gesture.

Figure 2: The co-speech gesture LARGE, accompanying the phrase "a bottle he liked" in (2).

While judgments are delicate, this sentence has been argued to give rise to three inferences: (i) that a philosopher found a bottle that he liked, (ii) that no other philosopher found a bottle that he liked, and (iii) that the bottle the philosopher found was large. Notice that the meaning of the gesture appears to contribute to modifying only the positive (i), namely that the philosopher found a *large* bottle that he liked, and not to the negative (ii) (that no other philosopher found a *large* bottle that he liked). Part of the puzzle raised by co-speech gestures, then, is to understand the apparently targeted nature of their contribution to the sentences that they modify.

1.2 Possible theories

While there is not yet a general consensus in the theoretical literature as to how gesture content is semantically related to linguistic content, we summarize here three main formal linguistic theories that one might consider.

1.2.1 At-issue analysis

One possibility is to take co-speech gestures to make an at-issue contribution to the meanings of the sentences they modify. Co-speech gestures would essentially be akin to modifiers such as "like this", where "this" refers to the relevant gesture. The gesture LARGE in (3a), for example, would be interpreted as though the *largeness* of the bottle had been explicitly asserted, along the lines of (3b). On this view, (3a) and (4a) should have the same meaning as (3b) and (4b), respectively.

(3)   a.   I brought [a bottle]_LARGE to the talk.
       b.   I brought a bottle that was [this]_LARGE large to the talk.

(4)   a.   John [helped]_LIFT his son.
       b.   John helped his son like [this]_LIFT.

An at-issue analysis of co-speech enrichments has not, to our knowledge, been seriously entertained in the literature, because co-speech enrichments appear to *project* out of the scope of various operators. For example, consider the contrast between (5a) and (5b):

(5)   a.   John didn't [help]_LIFT his son.
       b.   John didn't help his son like [this]_LIFT.

Unlike (5a), (5b) triggers the (defeasible) implicature that John helped his son. This is because "John didn't help his son like [this]_LIFT" evokes the more informative alternative "John didn't help his son", hence the implicature that John did help his son. On the other hand, (5a) arguably carries two inferences: first, that John didn't help his son *at all* (upward or otherwise), and second, that if he *had* helped his son, it would have been in an upward direction. But if co-speech enrichments are treated in the same way as "like this" modifiers, they should be expected to trigger the same implicatures. An at-issue analysis therefore needs to account for why (5a) and (5b) lead to distinct inferences.

1.2.2 Supplemental analysis

Because co-speech enrichments do not interact with logical operators in the same way as at-issue material, Ebert & Ebert (2014) take co-speech gestures to make a supplemental contribution, i.e. the same kind of contribution that appositive relative clauses make. For Ebert and Ebert, a sentence like (6a), for example, in which the gesture in Figure 2 aligns with "a bottle", would mean something like (6b). Specifically, the size of the bottle is not at issue (Ebert & Ebert 2014).

(6)   a.   I brought [a bottle]_LARGE to the talk.
        b.   I brought a bottle, which (by the way) was [this]_LARGE large, to the talk.

It is worth noting that this analysis also accounts for the more subtle inferences in (2); in particular, it captures the observation that the gestural inference clearly enriches the positive part of the asserted content (*a philosopher found a bottle he liked*), but not its negative component (*no other philosopher found a bottle that he liked*).

   By similar reasoning, (4a), repeated below as (7), could be given an analysis akin to (8a) or to (8b).

(7)   John [helped]_LIFT his son.

(8)   a.   John helped his son, which happened like [this]_LIFT.
        b.   John helped his son, which involved doing [this]_LIFT.

Because of the complexity of the behavior (and analysis) of appositives, there are several choice points in the Supplemental theory. For present purposes, the main tenet of the theory should be that a co-speech gesture gives rise to readings that can be paraphrased with appositive relative clauses. But one can obtain different versions of the theory depending on how liberal one is (i) with respect to the *size* of the antecedent of the appositive (corresponding to a verb phrase (VP) or to a full clause), and also (ii) with respect to the *mood* of the supplemental paraphrase (which may be indicative or subjunctive); of course, some choices make more sense than others.

   Consider the first choice point regarding the size of the appositive antecedent. In the main cases under study, the co-speech gesture co-occurs with the VP, and thus it is reasonable to assume that a sentence like (7) can be analyzed along the lines of (8b), where *which* refers back to (actions denoted by) the VP, rather than (8a), where *which* refers back to (events denoted by) the entire clause. No obvious difference arises in this particular case, but in some quantificational cases the difference may matter. In (9a), for instance, the appositive only provides information about the guy who helped his son, whereas (9b) may allow for a stronger inference to the effect that for each of the ten guys, helping one's son involved lifting him.

(9)   a.   Exactly one of these ten guys helped his son, which happened like [this]_LIFT.
        b.   Exactly one of these ten guys helped his son, which involved doing [this]_LIFT.

Concerning the second choice point regarding the mood of the supplemental paraphrase, different predictions are made if the appositive targets the entire clause, but possibly not if it targets just the VP. To make the point concrete, consider (10), which could be analyzed as (11) with an indicative or as (12) with a subjunctive. In the (a) examples, the gesture modifies the clause (i.e. *The event of John's helping his son (would have) happened with upwards lifting*), while in the (b) examples, the gesture modifies the VP (i.e. *The action of helping his son (would have) involved upwards lifting*). The difference in predictions is clear for the analysis on which the gesture modifies a full clause (notice the asymmetry between (11a) and (12a)), but not for the more plausible analysis on which the gesture modifies just the VP (i.e. (11b), (12b)).

(10)   John didn't [help]_LIFT his son.

(11)   a.   ?John didn't help his son, which happened like [this]_LIFT.
        b.   John didn't help his son, which involved doing [this]_LIFT.

(12)   a.   John didn't help his son, which would have happened like [this]_LIFT.
        b.   John didn't help his son, which would have involved doing [this]_LIFT.

At this point, then, it seems reasonable to focus on a version of the Supplemental analysis on which the gestural supplement modifies the VP – a situation in which the mood of the appositive does not matter in any obvious way. Be that as it may, it is worth pointing out two properties that are shared by all versions of the supplemental analysis discussed above. First, in the cases at hand, it is very difficult to understand the supplement as making an at-issue contribution within the scope of an operator, as in (9) and (10). This will matter when we compare the Supplemental analysis to the Cosuppositional analysis discussed below, which has a natural mechanism of *local accommodation* that yields precisely the relevant readings.[1] Second, depending on the size of the antecedent of the non-restrictive pronoun, and on the mood of the appositive, one may obtain inferences about all of the subject NP agents, or about all of the subject NP agents *that satisfy the VP*. But one cannot obtain further readings, and in particular not existential ones on which one infers that at least one of the subject agents should satisfy the supplemental condition. To be concrete, consider (13). In these cases, we obtain an inference that for *each* of the relevant guys, helping his son involved lifting, and we certainly don't get a reading on which the requirement is only that for at least *some* of these ten guys this condition was satisfied.

(13)   a.   None of these ten guys helped his son, which would have happened like
              [this]_LIFT.
        b.   None of these ten guys helped his son, which involved/which would have
              involved doing [this]_LIFT.

For further discussion of possible choice points in the Supplemental analysis, see Schlenker (To appear a; b).

1.2.3 Cosuppositional analysis

Running counter to the predictions of the Supplemental analysis, Schlenker (To appear a; b) observes that some environments support the presence of co-speech gestures, while appearing to disallow their appositive counterparts. The sentence in (14a), for example, seems to be acceptable while its appositive counterpart (14b) is not. Furthermore, if (14a) triggers any inference at all, it appears to be the one in (14c), which does not follow in any obvious way from (14b) (even if it were acceptable). Following a suggestion by Miloje Despic (p.c.), one can include "by the way" to force an appositive reading of a relative clause that might otherwise be read as being restrictive.

(14)   a.   No philosopher brought [a bottle of water]_LARGE to the talk.
        b.   #No philosopher brought a bottle of water, which (by the way) was
              [this]_LARGE large, to the talk.
        c.   ?⤳ *If a philosopher were to bring a bottle of water to the talk, it would be*
              [this]_LARGE large.

The debate is complicated by the fact that the supplement could be assumed to take an invisible subjunctive mood, as briefly mentioned above. One important part of the argument in Schlenker (To appear a; b) is that certain kinds of gestures, namely *post-speech* gestures that come *after* the expressions they modify rather than co-occurring with them,

---

[1] Local accommodation is the process by which a presupposition can, at some cost, behave as if it were part of the at-issue or assertive component.

do appear to give rise to supplement-like behavior. While we cannot go into the details of the argument here, its initial plausibility can be illustrated by the similarity between the post-speech gesture and indicative mood appositive examples in (15).

(15)   a.   A philosopher brought a bottle of water – LARGE.
        b.   A philosopher brought a bottle of water, which (by the way) was
              [this]_LARGE large.
        c.   ?No philosopher brought a bottle of water – LARGE.
        d.   ?No philosopher brought a bottle of water, which (by the way) was
              [this]_LARGE large.

In order to capture the acceptability of co-speech gestures in all environments, by contrast with indicative appositives and post-speech gestures, Schlenker (To appear a; b) proposes that co-speech gestures trigger presuppositions, and more specifically, *conditionalized presuppositions* (or *cosuppositions*). Like the presuppositions triggered by spoken phrases such as those in (16) and (17), which have been shown to project universally from "none"-NP (Chemla 2009), the inferences of co-speech gestures like SLAP (Figure 3) in (18) should then also show the same projection behavior. Importantly, in (18), the inference is tantamount to: *for each of these ten guys, if he were to punish his son, he would do so by slapping him* – which makes clear the conditionalized nature of the presupposition.

(16)   None of my students knew that he was incompetent.
        ↝ *Each of my students was incompetent (and male).*

(17)   None of these ten students takes good care of his computer.
        ↝ *Each of these ten students has a computer (and is male).*

(18)   None of these ten guys will [punish]_SLAP his son.
        ↝ *Each of these ten guys would punish his son by slapping him.*

Schlenker (To appear a; b) formalizes these intuitions within a dynamic semantics (see Heim 1983; Schlenker 2009), according to which presuppositions must be satisfied in their local contexts. That is, they must be entailed by the local contexts of the expressions that trigger them. Co-speech gestures, then, trigger presuppositions that their content is entailed by that of the expressions they modify:

Figure 3: The co-speech gesture SLAP, accompanying the verb "punish" in (18).

(19)   **Cosuppositions triggered by co-speech gestures**
        Let G be a co-speech gesture co-occurring with an expression d, and let g be the
        content of G. Then G triggers a presupposition $d \Rightarrow g$, where $\Rightarrow$ is generalized
        entailment (among expressions whose type ends in t).

The presuppositions triggered by co-speech gestures are thus special in that they are
conditionalized on the assertive content of the expressions they co-occur with. Such a
view of co-speech gestures predicts that the inferences they trigger will, much like ver-
bal presuppositions, project out of various linguistic environments, including questions,
negation, and quantifiers. One key question is therefore how presuppositions project from
quantified structures. Here there are two main theories to consider (though note that the
experimental results discussed in Chemla 2009 argue for a more nuanced view, namely
one that matches one theory or the other depending on the nature of the quantifier under
consideration).

On the Universal Projection theory, propounded by Heim (1983) and Schlenker (2009),
among others, all quantifiers trigger a universal presupposition or something close to
it. Concretely, an example such as (16) is predicted by such theories to yield the infer-
ence that *each* of the relevant ten students was incompetent. When combined with the
Cosuppositional analysis of co-speech gestures, these theories predict that (18) should
trigger the inference that for *each* of the relevant ten guys, if he were to punish his son,
he would do so by slapping him.

On the Existential Projection theory, put forth by Beaver (2001), presuppositions pro-
ject existentially from quantified structures. On this view, (16) is predicted to trigger the
inference that *at least one* of the relevant ten students was incompetent. The Existential
Projection version of the Cosuppositional analysis therefore likewise predicts that (18)
should trigger the inference that for *at least one* of the relevant ten guys, if he were to pun-
ish his son, he would do so by slapping him.

As noted above, some choice points in the Supplemental analysis make it possible for a
version of it to predict *universal* inferences, in a way that comes very close to the Universal
Projection theory. Specifically, by combining a liberal version of the Supplemental analy-
sis with the claim that the gesture-*qua*-appositive modifies the VP, we can obtain some
sort of universal inference in quantificational cases, along the lines of (13). On the other
hand, no plausible version of the Supplemental analysis comes close to predicting *existen-
tial* projection along the lines of the Existential Projection theory. A finding of such pat-
terns of projection would therefore provide an argument against the Supplemental view
of co-speech gestures.

Aside from the introspective judgments reported in Ebert & Ebert (2014) and Schlenker
(To appear a; b), there have been no experimental investigations of the ways in which
co-speech gestures interact with the logical structure of the sentences in which they are
found. As we will see, while our experimental results are far from definitive, they pose
definite problems for several theories: they rule out the At-issue theory, they raise prob-
lems for (versions of) the Supplemental theory and for the Universal Projection version of
the Cosuppositional theory, and they are more compatible with the Existential Projection
version of the Cosuppositional theory.

1.3 Experimental investigation

Previous works have investigated various aspects of the production, perception, process-
ing, and development of gestures (e.g., Kelly & Church 1998; Kelly & Barr 1999; Mayberry
& Nicoladis 2000; McNeil et al. 2000; O'Neill et al. 2002; Holler & Beattie 2003a; b; Holle
& Gunter 2007; Özyürek et al. 2007; Alibali et al. 2009; Gullberg 2009; Kelly et al. 2009;

Kidd & Holler 2009; Botting et al. 2010; Göksun et al. 2010; Cartmill et al. 2012; Dick et al. 2012; Lücking et al. 2012; Özçalişkan & Dimitrova 2013; Emmorey & Özyürek 2014; Hrabic et al. 2014; Özyürek 2014; Wagner et al. 2014). While many of these existing studies have examined the meanings that co-speech gestures contribute, they do not target the precise ways in which gestures may interact with the logical structure of the sentences with which they co-occur. In order to more precisely investigate the inference projection properties of co-speech gestures, then, we turn next to our experiments, designed to detect distinct interpretation strategies associated with iconic directional co-speech gestures.

2 Experimental design features

Our goal is to establish the possible interpretation strategies associated with co-speech gestures. Depending on the theory, a co-speech gesture may or may not give rise to local accommodation, existential projection, etc. In various linguistic environments such as negation and quantified sentences, these interpretation strategies correspond to specific readings. We tested the availability of these readings in two experiments, one using a Truth Value Judgment Task (Crain & Thornton 1998) and another using a Picture Selection Task. In both of these experiments, participants were asked to judge whether various sentences involving co-speech gestures matched the accompanying images, where the images made various relevant readings true or false. Sections 3 and 4 provide the details of the experiments and the results. Before moving to these, we first present the common features of the two experiments, which will then serve as a reference point for our discussion in Sections 3 and 4.

2.1 Sentences: 2 gestures, 6 linguistic environments, 2 conditions

To systematically test for the inferences of co-speech gestures, we will focus our attention on a specific pair of gestures, namely the directional gestures UP and DOWN. Figure 4 provides a screenshot of the co-speech gesture UP, produced with the index finger pointed upwards.

To investigate their projection properties, we will examine the interpretation of these directional gestures in six different linguistic environments: plain affirmative sentences (UNEMBEDDED), negative sentences (NEGATION), modal sentences (MIGHT), and quantified sentences (EACH, NONE, and EXACTLY-ONE), as in (20) through (25), respectively (see Appendix B for the complete list of sentences):

Figure 4: Screenshot of the co-speech gesture UP, which aligned either with "use the stairs" in the GESTURE condition (see (26)) or with "in this direction" in the ASSERTED condition (see (27)). The arrow indicates the upwards motion of the gesture.

(20)   The boy will [use the stairs]_UP.

(21)   The boy will not [use the stairs]_UP.

(22)   The boy might [use the stairs]_UP.

(23)   Each of these three boys will [use the stairs]_UP.

(24)   None of these three boys will [use the stairs]_UP.

(25)   Exactly one of these three boys will [use the stairs]_UP.

We will compare the interpretation of the directional gestures in target sentences, where the direction is merely gestured (26), with controls where the gesture is supported by the verbally asserted phrase "in this direction" (27).[2] If a particular projection pattern or interpretive strategy is specific to the gesture, it should not depend on the support of the verbally asserted phrase. Therefore, we can more confidently conclude that a projection pattern is contributed by the gesture if the pattern arises *more* in the GESTURE condition than in the ASSERTED condition.[3]

(26)   The boy will [use the stairs]_UP.

(27)   The boy will use the stairs [in this direction]_UP.

2.2 Contexts and images

To test for the possible semantic contributions of directional gestures, we have created contexts in which cartoon characters can use the stairs either to go up or to go down. A character who appears at the bottom of the stairs can only go *up* the stairs, while a character at the top of the stairs can only go *down* the stairs. Because the characters only ever appear at the top or the bottom of the stairs, it is clear that they can only go in one of the two directions. This will allow us to precisely pinpoint the direction as either *up* or *down*, with the target gestures being either compatible or incompatible with the visually depicted context.

   Being able to depict an upwards use of the stairs versus a downwards use of the stairs is not sufficient, however. Because the inferences of interest are *conditionalized*, some of the

---

[2] We take the semantics of the control sentence to be uncontroversial: "this" is a deictic expression whose denotation must be provided by the context, the gesture serves to make a certain manner of action salient in the context, and "like this" is an at-issue modifier. We will refer to the sentence containing "like this" as an ASSERTED control.

[3] We will end up departing somewhat from this conservative rule of interpretation, due to constraints inherent to the kinds of gestural sentences we are testing. In particular, it is possible that a projection pattern is available for the GESTURE sentence, and yet our test does not detect that it is more present there than in the ASSERTED control. For example, notice that the local accommodation reading is actually the *only* reading that participants should access in the ASSERTED condition; we should therefore never expect to find *more* of it in the GESTURE condition. Yet if we observe *some* amount of local accommodation in the GESTURE condition, we might want to conclude that it is a genuinely possible reading, even if there is less of it in the GESTURE condition than in the ASSERTED condition. Notice also that specifically in our experiments, the ASSERTED targets could easily become equivalent to the GESTURE targets merely by ignoring the "in this direction" phrase. If for whatever reason this were to happen, readings genuinely associated with the GESTURE condition might also be detected to some extent in the ASSERTED condition (though this would raise red flags in cases where no theory predicts such readings for the ASSSERTED targets). Our point here is simply that the difference that we aim to measure between the two conditions may be an underestimation.

contexts must be compatible with a *hypothetical* use of the stairs in a particular direction. In these cases, the character will crucially be blocked from using the stairs (i.e. by a barrier), despite appearing either at the top or the bottom of the stairs. This creates the possibility of a conditional inference: if the character were to use the stairs, s/he would clearly have to go in only one of the two possible directions. Restricting the possibilities in this way will enable us to systematically create the necessary contexts to test for the presence of the cosuppositional inferences.[4]

2.3 Combinations of sentences and images

Our goal is to investigate how participants treat the meanings that are conveyed by co-speech gestures. We will neutrally refer to the meanings contributed by the directional gestures UP and DOWN as *directional inferences*. For a given sentence, we have designed target images that are compatible with different directional inferences or, to put it differently, with different interpretation strategies: *Ignore directional inference*, *Existentially project directional inference*, *Universally project directional inference*, and *Locally accommodate directional inference*. Of course, not all interpretation strategies are meaningful for a given linguistic environment; existential and universal projection, for example, are not applicable in the non-quantified environments. For the six environments, then, we have designed images that are compatible with all logically possible combinations of strategies.

The details for each of the six environments under investigation are provided in Appendix C; here, we illustrate the situation with the environment EACH. The sentence in (28) may be interpreted as in the paraphrases in (29), depending on whether participants *ignore* the contribution of the directional gesture, *existentially project* the directional inference from under the quantifier "each", or *universally project* the inference (indistinguishable in this case from *local accommodation*).

(28)   Each of these three girls will [use the stairs]_UP.

(29)   a.   **Ignore directional inference:** Each of the girls will use the stairs.
        b.   **Existentially project directional inference:** Each of the girls will use the stairs, and for at least one of the girls, if she uses the stairs it will be in an upwards direction.
        c.   **Universally project directional inference:** Each of the girls will use the stairs, and for each of the girls, if she uses the stairs it will be in an upwards direction.

These possible readings stand in an entailment relation, such that it is not possible for (29c) to be true without (29a) and (29b) also being true. Here, and for other contexts as well, we created images which would exemplify all logically possible combinations of readings. For EACH, this leads us to the images in Figure 5. Table 1 provides the expected

---

[4] During piloting, it appeared that some participants might respond as though the relevant inferences were about the characters *trying* or *wanting* to use the stairs in a particular direction, perhaps because the characters were facing the direction they presumably planned to go in. To avoid any confusion that the characters' intentions might introduce, we subsequently portrayed the *blocked* characters who would not use the stairs as engaging in some other activity, and indicated in the instructions to participants that some characters would be engaging in some other activity (juggling) rather than using the stairs.
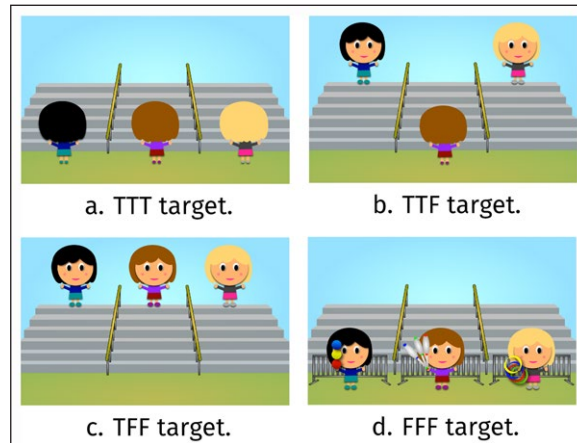
Figure 5: EACH target images accompanying the description "Each of these three girls will [use the stairs]_UP"/"Each of these three girls will use the stairs [in this direction]_UP". The TTT target was true on all readings; the TTF target was false only on the Universally Project reading; the TFF target was false on both the Existentially Project and Universally Project readings; the FFF target was false on all readings.

Table 1: Possible interpretation strategies in the EACH environment, and the corresponding truth values for the target sentences, when accompanied by each of the target pictures.

Interpretation strategy	Target pictures			
	TTT	TTF	TFF	FFF
Ignore directional inference	1	1	1	0
Existentially project directional inference	1	1	0	0
Universally project directional inference	1	0	0	0

truth values for the target sentence when accompanied by each of these target pictures, according to each of the possible interpretation strategies.

2.4 Additional controls (NO-GESTURE and NON-PATH)

Given our focus on directional gestures specifically involving the predicate "use the stairs", one might want to ensure that the predicate itself is not inherently associated with a bias for one particular direction, for example "using the stairs to go up". To determine whether such a bias exists, we will include NO-GESTURE controls. One control image is such that the character in question is at the bottom of the stairs, and another has the character at the top of the stairs (Figure 6). Both images will be accompanied by a description that is produced without a co-speech gesture. Crucially, since the description does not mention direction, and is therefore equally true of both images, any difference in the acceptance rates of the two control trials indicates an inherent directionality bias for the predicate "use the stairs".

One might also worry about the generalizability of the findings that we obtain from these specific directional pointing gestures UP/DOWN to other co-speech gestures. Restricting our attention to UP/DOWN allows us to focus on the specific readings of

interest, in a range of linguistic environments, in a systematic way. However, to ensure that participants are indeed sensitive to co-speech gestures beyond UP/DOWN, we will also include some clearly true and clearly false sentences that contain gestures describing manner rather than path. These NON-PATH controls involve characters going up or down in different ways, i.e. taking the stairs, using a slide, using a ladder, and using a rope. The images in these cases are accompanied by descriptions in which the speaker utters the direction (e.g., "The boy will go down") accompanied by a gesture indicating the manner of movement. An example is provided in Figure 7.

2.5 Analyses

Our goal is to decide whether there is evidence for a variety of interpretation strategies, such as existential projection and local accommodation. To do so, we use a version of the reading detection analysis described in Cremers & Chemla (2017). In essence, we model participants' responses using the different interpretation strategies as predictors. For instance, *existential projection* would be a predictor. Concretely, it would be a factor assigning value 1 to images in which this strategy predicts a true reading, and value 0 to images in which it predicts a false reading (see the corresponding line in Table 1).[5] If participants give true responses in some of the true conditions of a predictor and false
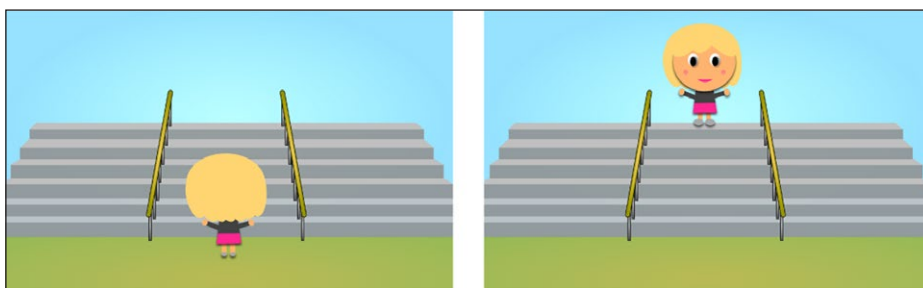
Figure 6: NO-GESTURE control images. The accompanying test sentence ("The girl will use the stairs") is produced without any gestures.
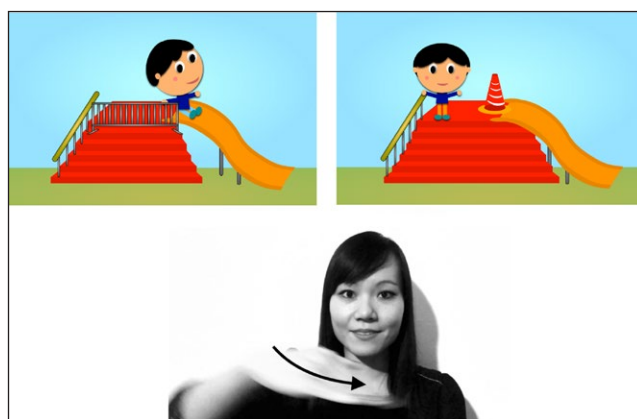

Figure 7: Two NON-PATH control images, accompanied by the description "The boy will [go down]_SLIDE", produced with a sliding gesture aligning with "go down". With the gesture, the description is a clearly true description of the image on the left, but a clearly false description of the image on the right.

---

[5] This description is most directly applicable to the Truth Value Judgment Task, but is easily translatable to the Picture Selection Task.

responses to the false conditions, this gives some weight to this predictor, i.e. it suggests that participants did use this interpretation strategy to some extent.[6] Note that in such an analysis the weight of a strategy is mitigated by the other strategies that happen to predict a true response in some of the same conditions. This is one of the advantages of the reading detection analysis: it quantifies the plausibility of a particular strategy, without ignoring the fact that other strategies may obscure the results if too few conditions are considered. This is critically important when there are more than two possible strategies that might be at play, as is the case here. In short, for each environment, we will obtain an estimate of the relative contribution of each interpretation strategy.

3 Experiment 1: Truth Value Judgment Task

We now present the results obtained from the Truth Value Judgment Task (TVJT), using the materials described in the previous section.

3.1 Method

3.1.1 Participants

Participants were recruited through Amazon Mechanical Turk, and were paid $1.20 for their participation. Two participants were excluded from analysis as they did not report English as (one of) their native language(s). Another 27 participants were excluded as they failed to score at least 70% accuracy on the NO-GESTURE and NON-PATH control trials (see Section 3.1.3 below). We report below the results from the remaining 172 participants (83 in the GESTURE condition and 89 in the ASSERTED condition).

3.1.2 Procedure

Participants were directed to a web-based TVJT, created and hosted on the Qualtrics platform. Participants saw a series of pictures depicting characters who appeared either at the top of the stairs, indicating they would use the stairs in a downwards direction, or at the bottom of the stairs, indicating they would use the stairs to go up. Participants saw one image at a time, accompanied by a video of one of the experimenters uttering a test sentence.[7] The participant's task was to decide whether the picture matched the speaker's description. Participants indicated their responses by clicking on "Yes" and "No" buttons. The task took about 10 minutes to complete. The instructions that participants saw are provided in Appendix A.1.

3.1.3 Materials

The details of the stimuli for each linguistic environment are provided in Appendix C. Condition (GESTURE vs. ASSERTED) was a between-subjects factor. In both conditions, participants saw two training items, followed by 34 test trials. Trial order was completely randomized across participants; subject NP gender (e.g., "the girl(s)"/"the boy(s)") and direction of the gesture (e.g., UP/DOWN) were also automatically randomized across trials. Participants saw all targets in the six different linguistic environments; all participants saw all linguistic environments.

In addition to the targets, participants saw 10 control trials. Two of the trials corresponded to the NO-GESTURE controls. These were meant to make sure that participants did not have an inherent bias to associate the predicate "use the stairs" with one specific

---

[6] More specifically, we will compare models with a particular strategy as a predictor and similar models without it. Evidence for the strategy being used in the population is obtained if the former model outperforms the latter.

[7] Each image appeared to the right of the video, so that the speaker's UP/DOWN gestures could not possibly be interpreted as pointing to (any aspects of) the test images.

direction. In addition to the NO-GESTURE controls, participants also saw four clearly true and four clearly false NON-PATH gesture controls.

3.2 Results

The data and R analysis script (R Core Team 2016) for this experiment are available online at http://semanticsarchive.net/Archive/GM0ZWNlM/Tieu-Pasternak-Schlenker-Chemla_Gestures.html. Here we will present the global results for the controls and targets. Details of the specific results from each linguistic environment can be found in Appendix D.

3.2.1 Controls

Only participants who correctly answered at least seven of the 10 control trials were included in the analysis. This criterion led to the inclusion of 172 participants in total (83 in the GESTURE condition, 89 in the ASSERTED condition). These participants' mean responses to the control trials are plotted in Figure 8. Given participants generally accepted both NO-GESTURE controls (where the character in question was at the top of the stairs and the bottom of the stairs, respectively), we can be reassured that participants did not have an inherent bias to associate *using the stairs* with one particular direction; that is, *using the stairs* could apply equally well to going up and going down the stairs.

3.2.2 Targets

Mean responses to the target conditions are presented in Figure 9. While the graph displays the average responses to each of the targets, Table 2 provides a summary of the *detectable* interpretation strategies. The presence or absence of the possible interpretation
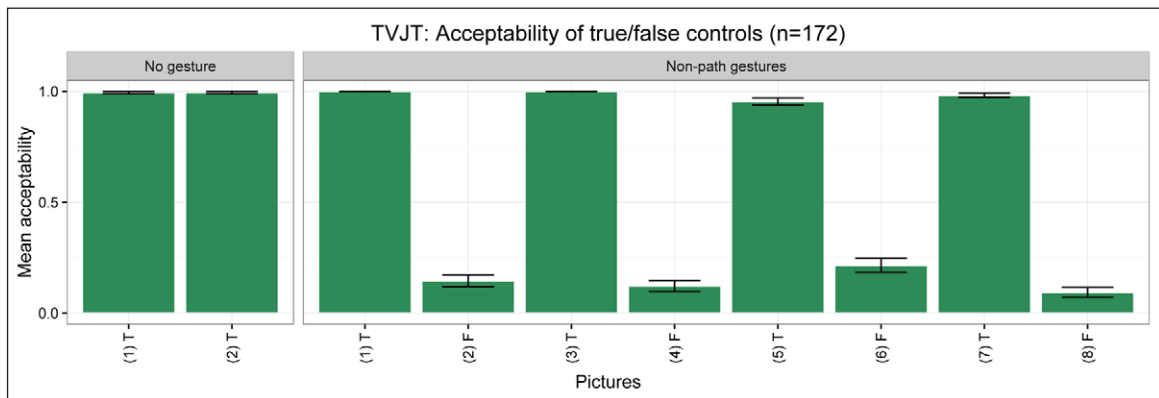

Figure 8: Mean acceptability of clearly true and clearly false NO-GESTURE and NON-PATH controls (target truth values are indicated in the picture names, along the x-axis).
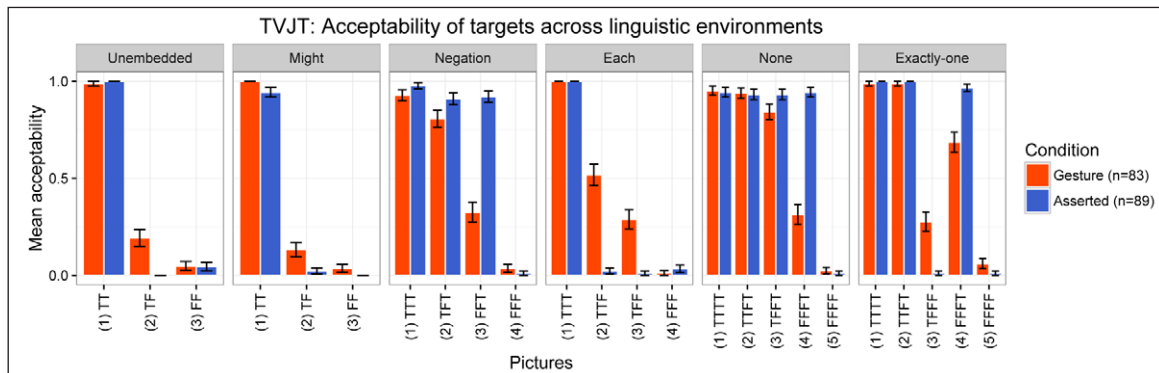

Figure 9: Mean acceptability of targets in each linguistic environment.

Table 2: Summary of the Truth Value Judgment Task results, indicating the availability of interpretation strategies in the GESTURE (GEST) and ASSERTED (ASRT) conditions. In certain cases, some readings were equivalent to *local accommodation* of the inference. Asterisks indicate cases where a strategy was found to be more available in one condition than in the mirror condition.

Environment	Interpretation strategies									
	Ignore		LocalAccom.		Project		Existential		Universal	
	GEST	ASRT	GEST	ASRT	GEST	ASRT	GEST	ASRT	GEST	ASRT
UNEMBEDDED	*			*						
MIGHT					(cf. LocalAccom.)					
NEGATION	*			*						
EACH	*								(cf. LocalAccom.)	
NONE	*			*						
EXACTLY-ONE				*						

	Tested and detected		Tested and not detected		Not tested/Not relevant

strategies was determined through a series of logistic regression models that were fitted to the GESTURE and ASSERTED responses in each linguistic environment. In Appendix D, we report on these models, which were run using the lme4 package in R (Bates et al. 2015; R Core Team 2016).[8]

As explained above (Section 2.5), we used a reading detection analysis as described in Cremers & Chemla (2017). Factors corresponding to each relevant interpretation strategy were defined (e.g., Ignore, Existential Projection and Universal Projection for the environment EACH), with value 1 for conditions in which the relevant interpretation was true and 0 when it was false (e.g., see the rows in Table 1). The types of models reported in the Appendix then attempt to predict responses by assigning optimal weights to each interpretation strategy, with a higher weight indicating that the corresponding interpretation strategy is more available.

In our analyses, we also evaluated more directly for each strategy whether it was *more* available in the GESTURE condition than in the ASSERTED condition, by examining the interaction between condition (GESTURE vs. ASSERTED) and the presence of the given strategy. This is important because if a strategy shows up as significant in both conditions, this does not rule out the possibility that the strategy is more available in one condition than in the other. It is also possible for a strategy to come out as significant in one condition but not the other, and yet the statistical evidence can fail to ensure that it really is *more* available in one condition than in the other (see Nieuwenhuis et al. 2011). In Table 2, we use asterisks to indicate the cases where a strategy was found to be more available in one condition than in the mirror condition. Importantly, note that in both the critical EACH and NONE conditions, the Existential Projection strategy does not turn out to be significantly more available in the GESTURE than in the ASSERTED condition, despite being detected in the former but not in the latter (see details in Appendix D). This could be due to a noisy estimation of the strategy in the ASSERTED condition, where we observe large confidence intervals. It could also be due to reasons discussed in Footnote 3. Hence, one should note that the evidence in favor of the Existential Projection strategy exists, but is weak. In addition, we see that Local Accommodation almost systematically comes out as more available in the ASSERTED condition. This does not show that Local Accommodation

---

[8] We attempted to fit logistic regression models on responses to all of the linguistic environments at once, with each of the interpretation strategies as fixed effects, but these models failed to converge.

is unavailable as a strategy for the GESTURE condition, as there is an intrinsic asymmetry here: local accommodation is the *only* interpretation that one should predict for the ASSERTED condition, while others are in principle available for the GESTURE condition (see, again, Footnote 3 for further discussion).

3.3 Discussion

We designed a Truth Value Judgment Task to detect for possible interpretation strategies of the directional co-speech gestures UP and DOWN. The experiment yields two main findings. First, in five of the six linguistic environments (MIGHT, NEGATION, EACH, NONE, EXACTLY-ONE), we observe that the directional inference in the GESTURE condition can be ignored, while the equivalent inference in the ASSERTED control is obligatorily integrated. Second, we observe existential projection of the directional inference from the scope of two quantifiers (EACH and NONE). Note that this projection pattern does not appear to be unique to the GESTURE condition (with the relevant interaction between Condition and Existential Projection being non-significant), but this lack of interaction may be due to a mere lack of power to process this analysis, in particular to properly evaluate the relevant (null) weight in the ASSERTED condition.

Let us now consider how these findings bear on the theories we outlined in Section 1.2. First, none of the theories *per se* predict that co-speech gestures can to some extent be disregarded. But this finding is not difficult to explain: co-speech gestures are easy to ignore because they are produced in a different modality from the spoken expressions they modify, and ignoring them still leaves us with a fully acceptable sentence. There is thus only a small penalty for ignoring them. By contrast, for the "like this" modifiers that served as our assertive controls, the gesture is needed to make sense of the modifier, otherwise one is faced with a demonstrative (= *this*) without a denotation. So at this point we take the Ignore possibility to be orthogonal to the main debate.

Second, the existence of projection effects in the target sentences may provide some evidence against the At-issue theory, although the results are not definitive, given that the observed projection behavior was not unique to the GESTURE condition. Additionally, that universal projection was mostly absent in our data would appear to argue against the Universal Projection version of the Cosuppositional theory. By contrast, the observed existential projection effects are compatible with the Existential Projection version of the Cosuppositional theory. The fact that local accommodation was observed is also in line with the Cosuppositional theory, at least on the view that co-speech gestures are weak presupposition triggers.

Third, one should discuss the import of these results for the Supplemental theories. Here, some specific items are worth discussing in detail. First, as seen in Appendix D.4, existential (but not universal) projection was observed in the EACH targets (although this did not interact significantly with Condition; that is, it was not more present in the GESTURE condition than in the ASSERTED condition). As we noted at the outset, as things stand, no plausible version of the Supplemental theory predicts existential projection. The point can be made concrete for embedding under "each" by considering (30): an appositive analysis would have to be developed along the lines of (31), which triggers a universal inference to the effect that for all of the relevant boys, using the stairs entails going up; a mere existential inference won't do.

(30)   Each of these three boys will [use the stairs]_UP.

(31)   Each of these three boys will [use the stairs], which they will do/which will involve going in [this]_UP direction.

In effect, the Supplemental theory encounters in this case the same problem as the Universal Projection version of the Cosuppositional theory. The difference between the two cases is that there is a plausible version of the Cosuppositional analysis with existential projection, whereas no similar empirical arguments have been made for existential projection of supplemental inferences.

A related difficulty involves the NONE targets, which also revealed evidence for existential projection (though again, the presence of Existential projection did not interact significantly with Condition). But as we already noted in connection with (13b), repeated below as (32), no version of the Supplemental theory can plausibly derive such an inference under "none". To reiterate the point using the examples at hand, the liberal version of the Supplemental analysis allows for an account of (33) along the lines of (34), but this only yields a universal inference.

(32)   None of these ten guys helped his son, which involved/which would have involved doing [this]_UP.

(33)   None of these three boys will [use the stairs]_UP.

(34)   None of these three boys will use the stairs, which they would do/which would involve going in [this]_UP direction.

An additional potential problem is that there is in this case some evidence for the possibility of local accommodation (though it was present to a greater degree in the ASSERTED condition). As Esipova (2016a; b) notes, it is currently unclear that supplements do give rise to local accommodation. One might want to argue that in the relevant cases the supplements have narrow scope, but this is usually taken to be a restricted option (e.g., Schlenker 2015), which is simply absent in the appositive controls that one might want to consider, such as (32) and (34).

Finally, the issue of local accommodation also arises for the EXACTLY-ONE targets (though here too participants were more inclined to locally accommodate on the ASSERTED targets). To be concrete, consider (35), evaluated against the FFFT picture in Figure 19. In this situation, it is simply *false* that *exactly one of these three boys will use the stairs* (since all three of them will); *a fortiori*, the sentence with the supplement (interpreted with wide scope) will be false as well. By contrast, it is *true* that *exactly one of these three boys will use the stairs in [this]_UP direction*. But the results indicate significant endorsement of the "true" response for the GESTURE target, which suggests that something akin to local accommodation is in fact available.

(35)   Exactly one of these three boys will [use the stairs]_UP.

(36)   Exactly one of these three boys will use the stairs, which he will do in [this]_UP direction.

Let us add that we also piloted a similar version of the experiment using a slider task rather than a TVJT, wherein participants could rate the acceptability of the sentences by dragging a slider to fill in a bar. The ends of the scale were indicated as "Not at all" (acceptable) to "Perfectly" (acceptable), and this was linearly mapped to a scale from 0–100 for analysis. Pilot results indicated that the slider experiment was not more informative about projective behavior than the TVJT, and moreover yielded less integration of

the gestures than the TVJT, so we decided not to run a full version of it. See Table 21 in the Appendix for a summary of the results of the slider pilot.

In assessing the efficacy of the TVJT methodology, we next considered the possibility that it may have been relatively easy for participants to *ignore* the gesture when the direction that it implied was not consistent with the target image. We thought that by pairing the relevant images side-by-side, we could more easily offer participants a chance to answer in a way that was consistent with the inference of the gesture. Of course, having to assess two images simultaneously, one consistent with the direction UP and the other with DOWN, for example, might make the task more difficult overall. On the other hand, we reasoned that it might also make the task easier by highlighting the respective directionality inferences. To investigate this possibility, we turned next to a Picture Selection Task.

4 Experiment 2: Picture Selection Task

The Picture Selection Task used the same materials as the TVJT, except that the critical target comparisons were presented directly to participants, two images at a time.

4.1 Method

4.1.1 Participants

Participants were recruited through Amazon Mechanical Turk, and were paid $1.20 for their participation. Four participants were excluded from analysis because they did not report English as (one of) their native language(s). We report below the results from the remaining 198 participants (99 in the GESTURE condition and 99 in the ASSERTED condition).

4.1.2 Procedure

Participants were directed to a web-based Picture Selection Task, created and hosted on the Qualtrics platform. As in the TVJT, participants saw pictures depicting characters who appeared either at the top of the stairs, indicating they would use the stairs in a downwards direction, or at the bottom of the stairs, indicating they would use the stairs to go up. Participants saw two images at a time, accompanied by a video of one of the experimenters uttering a test sentence. The participant's task was to decide which picture best matched the speaker's description. Participants indicated their responses by clicking on the picture of their choice. The task took about 10 minutes to complete. The instructions that participants saw are provided in Appendix A.2.

4.1.3 Materials

The details of the stimuli for each linguistic environment are provided in Appendix C. As in the TVJT, we presented test sentences containing directional descriptions involving six different linguistic environments. Again, each description contained a directional gesture; in the GESTURE condition, the direction was merely gestured, while in the ASSERTED condition, the gesture was supported by the verbally asserted phrase "in this direction". As in the TVJT, condition (GESTURE vs. ASSERTED) was a between-subjects factor, and linguistic environment was a within-subjects factor. In both the GESTURE and ASSERTED conditions, participants saw two training items followed by 20 test trials. The materials were the same as in the TVJT, except that the images that had appeared individually in the TVJT were paired. In some cases, the pairing was predicted to lead to a clear preference for one picture over the other, while other pairings were such that either both or neither of the pictures were good matches for the description.[9] The left-right order of the two

---

[9] Each pair of images was center-aligned beneath the video of the speaker, so that the speaker's UP/DOWN gestures could not possibly be interpreted as pointing to one of the two pictures.

target pictures being compared on any given trial was automatically randomized, as was trial order across participants. As in the TVJT, subject NP gender (i.e. "boy(s)"/"girl(s)") and direction of the gesture (i.e. UP/DOWN) were also randomized.

Instead of crossing the images from the TVJT to create all possible pairs, we selected for each environment a subset of pairwise comparisons from Section 2.3 that would allow us to evaluate the contributions of the interpretations of interest. In the case of the TVJT task from Experiment 1, it was crucial to evaluate all possible interpretations at once: the weight assigned to an interpretation strategy was calculated by factoring out the possible contribution of all the other possible strategies. Here, the evaluation is more direct and in some cases our choice of pairwise comparisons would be reduced and would not allow for an evaluation of the Ignore strategy; this was the case for NEGATION, NONE, and EXACTLY-ONE. The Ignore interpretation, however, is of little interest for the projection problem under investigation, and so dropping it from the analysis does not affect our claims about the availability of the other strategies. The pairings and predictions selected for each environment are given in full in Appendix E.

4.2 Results

The data and R analysis script for this experiment are available online at http://semantic-sarchive.net/Archive/GM0ZWNlM/Tieu-Pasternak-Schlenker-Chemla_Gestures.html. As before, we present here the global results for the controls and targets, the specific results from each linguistic environment can be found in Appendix F.

4.2.1 Controls

Mean responses to the control trials are plotted in Figure 10. Given that participants displayed chance performance on the NO-GESTURE control (where the two images corresponded to the character in question at the top of the stairs and the bottom of the stairs, respectively), we can be reassured that participants did not have an inherent bias to associate *using the stairs* with a particular direction; that is, *using the stairs* could apply equally well to going up and going down the stairs.
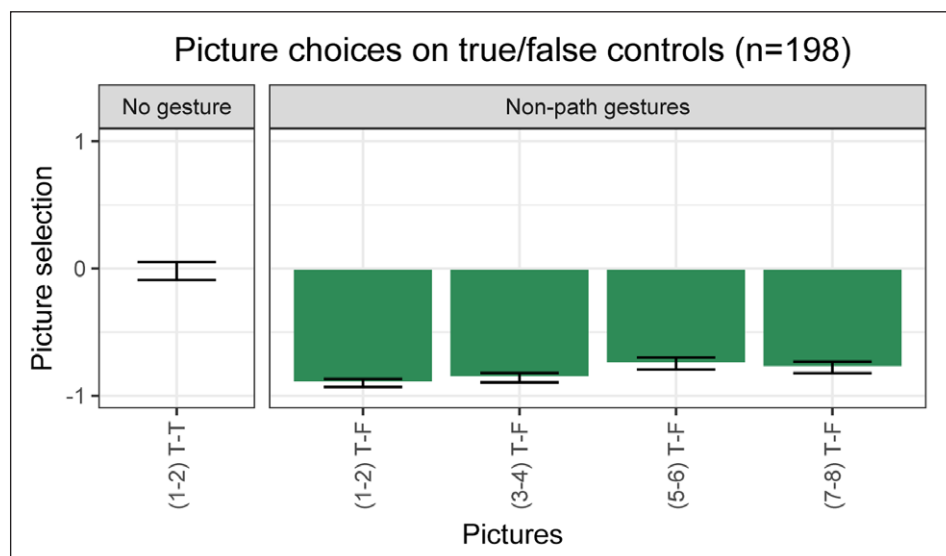

Figure 10: Rates of picture selections on NO-GESTURE and NON-PATH controls. The two pictures contrasted on each trial (e.g., (T)rue vs. (F)alse) are indicated along the x-axis; for easy visualization, selections of the left-labeled picture are coded here as –1, and selections of the right-labeled picture are coded as +1.

4.2.2 Targets

Mean responses to the target conditions are presented in Figure 11. A summary of the detectable interpretation strategies in the Picture Selection Task is provided in Table 3.

In Appendix F, we report on the logistic regression models we fitted to the data in the GESTURE and ASSERTED conditions, in each linguistic environment, in order to determine the presence or absence of the possible interpretation strategies (using the lme4 package in R, Bates et al. 2015; R Core Team 2016).[10] In all cases, selection of the left-labeled picture (generally the one with more true readings) was coded as –1, while selection of the right-labeled picture (generally the one with fewer true readings)
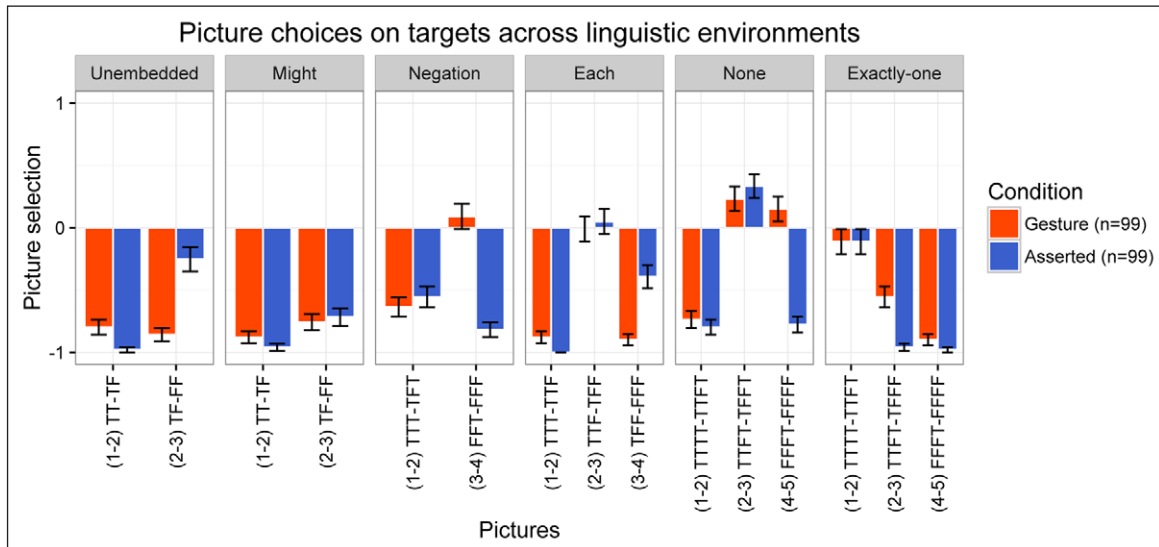
Figure 11: Rates of picture selections on the targets from each linguistic environment. The two pictures contrasted on each trial are indicated along the x-axis; for easy visualization, selections of the left-labeled picture are coded here as –1, and selections of the right-labeled picture are coded as +1.

Table 3: Summary of the Picture Selection Task results, indicating the availability of interpretation strategies in the GESTURE (GEST) and ASSERTED (ASRT) conditions.

Environment	Interpretation strategies									
	Ignore		**LocalAccom.**		**Project**		**Existential**		**Universal**	
	GEST	**ASRT**	**GEST**	**ASRT**	**GEST**	**ASRT**	**GEST**	**ASRT**	**GEST**	**ASRT**
UNEMBEDDED										
MIGHT					(cf. LocalAccom.)					
NEGATION										
EACH									(cf. LocalAccom.)	
NONE										
EXACTLY-ONE										

	Tested and detected		Tested and not detected		Not tested

---

[10] Unlike the TVJT, we do not report on combined GESTURE + ASSERTED models, which did not converge. We will therefore focus only on the individual models for each condition, in each linguistic environment. In each case, inspection of the reported confidence intervals helps us to evaluate the relative availability of a given interpretation strategy across the GESTURE and the ASSERTED conditions.

was coded as +1. Note that the left/right labeling merely reflects an internal coding scheme, and that the actual side of presentation of the pictures was randomized. What is crucial, then, is the alignment of this coding with the coding of the interpretation strategies.

Each of the possible interpretation strategies was modeled as fixed effects with three possible levels: –1, corresponding to a predicted selection of the left-labeled picture; +1, corresponding to a predicted selection of the right-labeled picture, and 0, corresponding to predicted chance performance (in cases where the target sentence was equally true or equally false of the paired images).

4.3 Discussion

On the whole, the Picture Selection Task appears to have detected fewer differences between the GESTURE targets and the ASSERTED controls than the TVJT experiment. In particular, the Picture Selection Task reveals no positive evidence for the existential projection pattern that was previously observed under "each". Additionally, however, we observe existential projection of the directional inference from "exactly one", the presence of which did not reach significance in the TVJT ($p = .07$). These differences in findings could be argued to demonstrate the non-robustness of the results, or they could be attributed to superficial differences between the two tasks that make each one more or less suited to different conditions. For instance, to detect subtle differences between two readings, it may be better to ask directly for an assessment of the contrast between two pictures that distinguish these readings (consider, for example, Figure 12). This would be close to what linguists actually do when creating and judging minimal pairs, thereby increasing the resolution of introspection (see Sprouse & Almeida 2012, as well as Marty et al. 2016 for similar arguments that contrastive judgments achieve higher experimental power). On the other hand, it could be that the picture selection task would be less suitable in other cases, where one reading might completely obscure another, despite both being available to some extent under appropriate conditions.

Another difference in the results from the two experiments is the degree to which participants were willing to ignore the directional phrase. We had reasoned that a picture selection task might make it easier for participants to identify those images that were consistent with inferences of the directional co-speech gestures, in contrast to those that were not. The results, however, suggest that perhaps the opposite was true: rather than highlighting the differences in directionality, seeing two images at a time in some cases may have encouraged participants to ignore the directional phrase. This was the case for MIGHT and EACH: the TVJT results revealed that the gesture could be ignored in
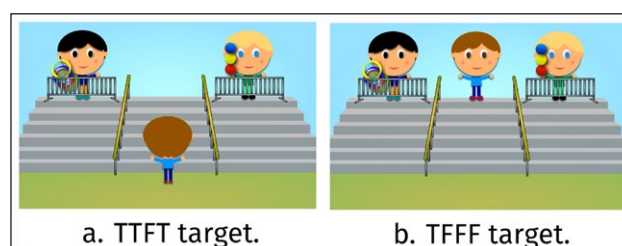

Figure 12: EXACTLY-ONE images accompanying "Exactly one of these three boys will [use the stairs]_UP". A participant who existentially projected the directional inference was expected to prefer the TTFT image over the TFFF image.

the GESTURE but not in the ASSERTED condition, whereas on the Picture Selection Task, participants appeared to ignore the gesture in both conditions. It may be that when the target images in Figure 13 were placed side by side, the importance of the directionality might somehow have been diminished, compared to when only a single image was presented at a time. More coarsely, it may be that the participants were more busy inspecting the *two* images in the picture selection task and paid less attention to the visual information present in the video.

Perhaps relatedly, on the Picture Selection Task, under NEGATION and NONE, the directional inference of the asserted control was locally accommodated while that of the gesture target was not. As seen in Figure 14, participants who locally accommodated the directional inference were expected to prefer the FFT/FFFT images over the FFF/FFFF images; the relevant images minimally differed by whether the characters were at the top or the bottom of the stairs. Again, it is possible that seeing pairs of images that differed only in directionality may have made it easier for participants to disregard the directional gesture.[11]

5 Conclusion

In this study, we used a Truth Value Judgment Task and a Picture Selection Task to investigate the projection properties of inferences arising from co-speech gestures in various linguistic environments. We began by summarizing the theoretical landscape, with three
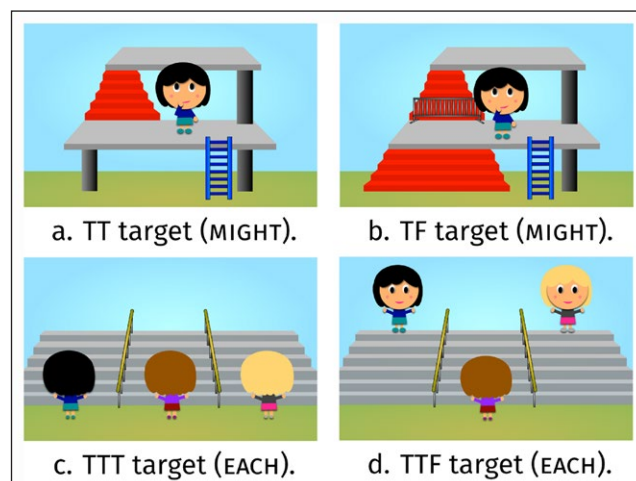
Figure 13: MIGHT and EACH images accompanying "The girl might [use the stairs]_UP"/"Each of these three girls will [use the stairs]_UP". A participant who projected the directional inference was expected to prefer the TT/TTT images over the TF/TTF images.

---

[11] As an anonymous reviewer points out, the *Ignore* strategy may in fact be meaningful, for instance about how easy or hard it is to integrate a co-speech gesture. For example, a participant might opt to ignore a gesture as a repair strategy for an otherwise ungrammatical or unacceptable sentence. We suspect that our experimental setting may have made it relatively easy and unproblematic for participants to ignore the gestures and still successfully complete the tasks; it is difficult, however, to infer that this would occur in natural face-to-face interactions. Note that if ignoring the strategy were the *only* strategy used to compensate for ungrammaticality, we would have a way to assess how strong this potential problem is. There is no reason to think that this is the only repair strategy, but we agree that it would be very useful in future studies to obtain accompanying acceptability judgments.
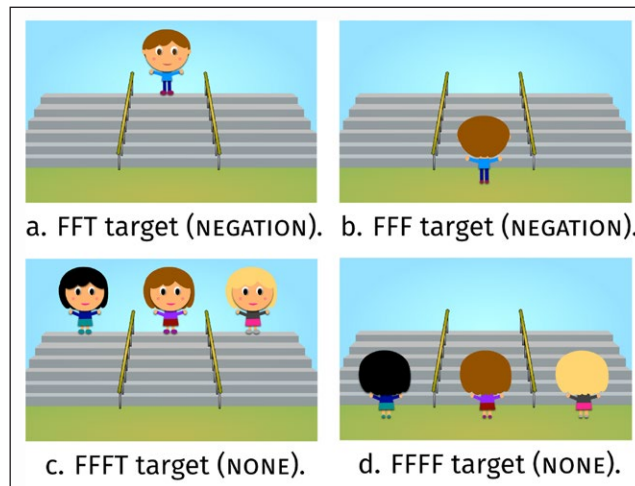
Figure 14: NEGATION and NONE images accompanying "The boy will not [use the stairs]_UP"/"None of these three girls will [use the stairs]_UP". A participant who locally accommodated the directional inference was expected to prefer the FFT/FFFT images over the FFF/FFFF images.

general theories that one might consider: the At-issue analysis, which takes co-speech gestures to make the same kind of enrichment as standard modifiers such as "like this"; the Supplemental analysis, which takes co-speech gestures to behave like appositive relative clauses; and the Cosuppositional analysis, which takes co-speech gestures to trigger presuppositions that are conditionalized on the contributions of the expressions they modify. We have observed some differences between the two tasks we used, which as mentioned may have to do with specific aspects of the tasks masking certain kinds of behavior. Nevertheless, taken together, the collective dataset leads to several conclusions.

First, all theories must be supplemented with the assumption that co-speech gestures can to some extent be disregarded. As mentioned above, this need not be surprising, since in our target sentences the co-speech gestures could be ignored without yielding an incoherent result, unlike the case of the "like this" controls. We would caution, however, that it is too early to tell whether the possibility of disregarding co-speech gestures is a robust finding, or merely a by-product of the experimental paradigms we selected.

Second, there is evidence of projection phenomena that are not predicted by the At-issue theory, but are more compatible with some version of the Cosuppositional theory.

Third, the present experiments yield evidence of existential but not universal projection from the scope of quantifiers, in particular under "each", "none", and "exactly one". This result can be explained by the Cosuppositional analysis, but only if it is combined with a theory of presupposition projection sometimes entertained in the literature, according to which presuppositions project existentially from the scope of quantifiers. Existential projection is very difficult to explain on the Supplemental theory. Still, it is worth noting the contrast between this finding and other experimental results indicating *universal* projection of presuppositions under the negative quantifier "no(ne)" (as reported in Chemla 2009, although see Zehr et al. 2015; 2016 for more recent discussion).

Finally, we have uncovered some evidence of *local accommodation* of the inferences of co-speech gestures (i.e. of partial *at-issue* behavior), which can be explained by the Cosuppositional theory but not necessarily by the Supplemental theory.

More generally, results of this kind further suggest commonalities and connections across the (verbal and visual) modalities, consistent with much previous work on gestures. Our results, however, suggest that the interaction between gesture and speech may be even deeper than previous treatments of gesture have assumed. In particular, while it is a common finding in the literature that gesture and speech both contribute to semantic processing, and that speakers rapidly integrate semantic information conveyed by gestures just as they do with spoken expressions, our experiments show that participants are in fact computing *inferences* from gestures, which interact in specific ways with the logical structure of their linguistic environments. Specifically, we find that participants can *project* the inferences of co-speech gestures from certain linguistic environments, just as they do with the presuppositions of verbal expressions. Future work might continue to explore the interplay between the two modalities, exploring a wider range of gestures (e.g., co-speech vs. post-speech gestures, Schlenker To appear a) and linguistic environments.

Additional Files

The additional files for this article can be found as follows:

- **A.** Instructions. DOI: https://doi.org/10.5334/gjgl.334.s1
- **B.** Test sentences. DOI: https://doi.org/10.5334/gjgl.334.s1
- **C.** Readings and relevant images for each linguistic environment. DOI: https://doi.org/10.5334/gjgl.334.s1
- **D.** Experiment 1: Results by environment. DOI: https://doi.org/10.5334/gjgl.334.s1
- **E.** Experiment 2: Pairings by environment. DOI: https://doi.org/10.5334/gjgl.334.s1
- **F.** Experiment 2: Results by environment. DOI: https://doi.org/10.5334/gjgl.334.s1
- **G.** Slider task. DOI: https://doi.org/10.5334/gjgl.334.s1

Competing Interests

The authors have no competing interests to declare.

References

Alibali, Martha W., Julia L. Evans, Autumn B. Hostetter, Kristin Ryan & Elina Mainela-Arnold. 2009. Gesture-speech integration in narrative. *Gesture* 9(3). 290–311. DOI: https://doi.org/10.1075/gest.9.3.02ali

Bates, Douglas, Martin Mächler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1). 1–48. DOI: https://doi.org/10.18637/jss.v067.i01

Beaver, David. 2001. *Presupposition and Assertion in Dynamic Semantics*. Stanford University, Stanford, CA: CSLI Publications.

Botting, Nicola, Nicholas Riches, Marguerite Gaynor & Gary Morgan. 2010. Gesture production and comprehension in children with specific language impairment. *British Journal of Developmental Psychology* 28(1). 51–69. DOI: https://doi.org/10.1348/026151009X482642

Cartmill, Erica A., Özlem Ece Demir & Susan Goldin-Meadow. 2012. Studying gesture. In Erika Hoff (ed.), *Research methods in child language: A practical guide, First edition*. Malden, MA/Oxford/West Sussex: Blackwell Publishing Ltd. DOI: https://doi.org/10.1002/9781444344035.ch14

Chemla, Emmanuel. 2009. Presuppositions of quantified sentences: Experimental data. *Natural Language Semantics* 17(4). 299–340. DOI: https://doi.org/10.1007/s11050-009-9043-9

Crain, Stephen & Rosalind Thornton. 1998. *Investigations in Universal Grammar: A guide to experiments on the acquisition of syntax and semantics*. Cambridge, Massachusetts: MIT Press.

Cremers, Alexandre & Emmanuel Chemla. 2017. Experiments on the acceptability and possible readings of questions embedded under emotive-factives. *Natural Language Semantics* 25(3). 223–261. DOI: https://doi.org/10.1007/s11050-017-9135-x

Dick, Anthony Steven, Susan Goldin-Meadow, Ana Solodkin & Steven L. Small. 2012. Gesture in the developing brain. *Developmental Science* 15(2). 165–180. DOI: https://doi.org/10.1111/j.1467-7687.2011.01100.x

Ebert, Cornelia & Christian Ebert. 2014. Gestures, demonstratives, and the attributive/referential distinction. Talk presented at the 7th Semantics and Philosophy in Europe, ZAS, Berlin, June 2014.

Emmorey, Karen & Asli Özyürek. 2014. Language in our hands: Neural underpinnings of sign language and co-speech gesture. In Michael S. Gazzaniga & George R. Mangun (eds.), *The cognitive neurosciences, Fifth edition*, 657–666. Cambridge, Massachusetts: MIT Press.

Esipova, Maria. 2016a. Alternatives matter: Contrastive focus and presupposition projection in standard triggers and co-speech gestures. Poster presented at MACSIM, CUNY, October 2016.

Esipova, Maria. 2016b. Presuppositions under contrastive focus: Standard triggers and co-speech gestures. Ms. New York University.

Göksun, Tilbe, Kathy Hirsh-Pasek & Roberta Michnick Golinkoff. 2010. How do preschoolers express cause in gesture and speech? *Cognitive Development* 25(1). 56–68. DOI: https://doi.org/10.1016/j.cogdev.2009.11.001

Gullberg, Marianne. 2009. Gestures and the development of semantic representations in first and second language acquisition. *Acquisition et interaction en langue étrangère... Languages, interaction, and acquisition (Aile...Lia)* 1. 117–139.

Heim, Irene. 1983. On the projection problem for presuppositions. In Daniel P. Flickinger (ed.), *Proceedings of the 2nd West Coast Conference on Formal Linguistics*, 114–125. Stanford University, Stanford, CA: CSLI Publications.

Holle, Henning & Thomas C. Gunter. 2007. The role of iconic gestures in speech disambiguation. *Journal of Cognitive Neuroscience* 19. 1175–1192. DOI: https://doi.org/10.1162/jocn.2007.19.7.1175

Holler, Judith & Geoffrey Beattie. 2003a. How iconic gestures and speech interact in the representation of meaning: Are both aspects really integral to the process? *Semiotica* 1(4). 81–116. DOI: https://doi.org/10.1515/semi.2003.083

Holler, Judith & Geoffrey Beattie. 2003b. Pragmatic aspects of representational gestures. *Gesture* 3(2). 127–154. DOI: https://doi.org/10.1075/gest.3.2.02hol

Hrabic, Melissa L., Rebecca A. Williamson & Şeyda Özçalişkan. 2014. How early do children understand iconic co-speech gestures conveying action? In Will Orman & Matthew James Valleau (eds.), *Online supplement to the Proceedings of the 38th Boston University Conference on Language Development.*

Kelly, Spencer D. & Dale J. Barr. 1999. Offering a hand to pragmatic understanding: The role of speech and gesture in comprehension and memory. *Journal of Memory and Language* 40. 577–592. DOI: https://doi.org/10.1006/jmla.1999.2634

Kelly, Spencer D., Peter Creigh & James Bartolotti. 2009. Integrating speech and iconic gestures in a Stroop-like task: Evidence for automatic processing. *Journal of Cognitive Neuroscience* 22(4). 683–694. DOI: https://doi.org/10.1162/jocn.2009.21254

Kelly, Spencer D. & R. Breckinridge Church. 1998. A comparison between children's and adults' ability to detect conceptual information conveyed through representational gestures. *Child Development* 69(1). 85–93. DOI: https://doi.org/10.1111/j.1467-8624.1998.tb06135.x

Kidd, Evan & Judith Holler. 2009. Children's use of gesture to resolve lexical ambiguity. *Developmental Science* 12(6). 903–913. DOI: https://doi.org/10.1111/j.1467-7687.2009.00830.x

Lücking, Andy, Kirsten Bergman, Florian Hahn, Stefan Kopp & Hannes Rieser. 2012. Data-based analysis of speech and gesture: the Bielefeld Speech and Gesture Alignment corpus (SaGA) and its applications. *Journal of Multimodal User Interfaces* 7(1–2). 5–18.

Marty, Paul, Jon Sprouse & Emmanuel Chemla. 2016. Evaluating data collection methods in linguistics: Modes of presentations and scales of judgments for grammaticality judgments. Ms., Massachusetts Institute of Technology.

Mayberry, Rachel I. & Elena Nicoladis. 2000. Gesture reflects language development: Evidence from bilingual children. *Current Directions in Psychological Science* 9(6). 192–196. DOI: https://doi.org/10.1111/1467-8721.00092

McNeil, Nicole M., Martha W. Alibali & Julia L. Evans. 2000. The role of gesture in children's comprehension of spoken language: Now they need it, now they don't. *Journal of Nonverbal Behavior* 24(2). 131–150. DOI: https://doi.org/10.1023/A:1006657929803

Nieuwenhuis, Sander, Birte U. Forstmann & Eric-Jan Wagenmakers. 2011. Erroneous analyses of interactions in neuroscience: a problem of significance. *Nature Neuroscience* 14. 1105–1107. DOI: https://doi.org/10.1038/nn.2886

O'Neill, Daniela K., Jane Topolovec & Wilma Stern-Cavalcante. 2002. Feeling sponginess: The importance of descriptive gestures in 2- and 3-year-old children's acquisition of adjectives. *Journal of Cognition and Development* 3(3). 243–277. DOI: https://doi.org/10.1207/S15327647JCD0303_1

Özçalişkan, Seyda & Nevena Dimitrova. 2013. How gesture input provides a helping hand to language development. *Seminars in Speech and Language* 34(4). 227–236. DOI: https://doi.org/10.1055/s-0033-1353447

Özyürek, Asli. 2014. Hearing and seeing meaning in speech and gesture: Insights from brain and behaviour. *Philosophical Transactions of the Royal Society B* 369. 20130296. DOI: https://doi.org/10.1098/rstb.2013.0296

Özyürek, Asli, Roel M. Willems, Sotaro Kita & Peter Hagoort. 2007. Online integration of semantic information from speech and gesture: Insights from event-related brain potentials. *Journal of Cognitive Neuroscience* 19. 605–616. DOI: https://doi.org/10.1162/jocn.2007.19.4.605

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 and Pragmatics* 2(3). 1–78. DOI: https://doi.org/10.3765/sp.2.3

Schlenker, Philippe. 2015. The semantics and pragmatics of appositives. Ms. Institut Jean-Nicod and New York University.

Schlenker, Philippe. To appear a. Iconic pragmatics. *Natural Language & Linguistic Theory*.

Schlenker, Philippe. To appear b. Gesture projection and cosuppositions. *Linguistics and Philosophy*.

Sprouse, Jon & Diogo Almeida. 2012. Power in acceptability judgment experiments and the reliability of data in syntax. Ms. University of California, Irvine.

Tieu, Lyn, Robert Pasternak, Philippe Schlenker & Emmanuel Chemla. 2016. Co-speech gesture projection: Evidence from inferential judgments. Ms. Ecole Normale Supérieure.

Wagner, Petra, Zofia Malisz & Stefan Kopp. 2014. Gesture and speech in interaction: An overview. *Speech Communication* 57. 209–232. DOI: https://doi.org/10.1016/j.specom.2013.09.008

Zehr, Jérémy, Cory Bill, Lyn Tieu, Jacopo Romoli & Florian Schwarz. 2015. Existential presupposition projection from *none*: An experimental investigation. In Thomas Brochhagen, Floris Roelofsen & Nadine Theiler (eds.), *Proceedings of the 20th Amsterdam Colloquium*, 448–457.

Zehr, Jérémy, Cory Bill, Lyn Tieu, Jacopo Romoli & Florian Schwarz. 2016. Presupposition projection from the scope of *None*: Universal, existential, or both? In Mary Moroney, Carol-Rose Little, Jacob Collard & Dan Burgdorf (eds.), *Proceedings of the 26th Semantics and Linguistic Theory Conference*, 754–774. DOI: https://doi.org/10.3765/salt.v26i0.3837