

# Acceptability of at-issue co-speech gestures under contrastive focus

Maria Esipova  
*New York University*  
10 Washington Place, New York, NY  
10003, USA  
[masha.esipova@nyu.edu](mailto:masha.esipova@nyu.edu)

**Abstract** The status of content-bearing co-speech gestures, i.e., gestural adjuncts co-occurring with the verbal expressions they adjoin to, has recently become a matter of debate in formal semantics and pragmatics (Ebert & Ebert 2014; Ebert 2017; Schlenker 2018; Tieu et al. 2017a; b; Esipova 2018; Zlogar & Davidson 2018). The general tendency has been to claim that co-speech gestures by default make not-at-issue contributions, however, the existing analyses differ in whether they in principle allow for at-issue interpretations of co-speech gestures and, if yes, in how much cost such at-issue interpretations can incur. In this study I use an acceptability judgement task to investigate the acceptability of at-issue interpretations of co-speech gestures forced by contrastive focus, as well as some factors that can potentially affect that acceptability. I conclude that while the overall results are in principle compatible with any analysis that posits a (strong) bias against at-issue interpretations of co-speech gestures, further inspection of individual variation in judgement patterns allows us to argue against analyses in which the level of such bias is fixed across speakers. In particular, the variation data can be taken as evidence against the analysis of co-speech gestures as Pottsian (2005) supplements akin to appositives (Ebert & Ebert 2014; Ebert 2017). As for the factors that can potentially affect the acceptability of at-issue interpretations of co-speech gestures under contrastive focus, neither the type of content encoded by the gesture, nor emphatic production of co-speech gestures have been found to have an effect.

**Keywords:** co-speech gestures, (not)-at-issueness, contrastive focus, acceptability judgements, variation

# 1 Introduction

## 1.1 Co-speech gestures: background

This paper focuses on *co-speech gestures*, i.e., content-bearing, non-conventionalized gestures that co-occur with some verbal expression and contribute some further information about its denotation:



- (1) John might order a beer<sup>LARGE</sup>.<sup>1</sup>

It has been claimed in the literature (Ebert & Ebert 2014; Schlenker 2018; Tieu et al. 2017a; b) that (1) gives rise to an inference that if John orders a beer, it will be large (i.e., John won't order a small beer).<sup>2</sup>

In contrast, (2), a counterpart of (1) with an adjectival modifier, doesn't give rise to such an inference:

- (2) John might order a large beer.

For that reason the authors above conclude that the contribution of co-speech gestures is typically *not-at-issue*, because their content *projects* from (i.e., is preserved in) a variety of embedding environments, including from under *might*.

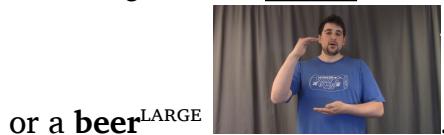
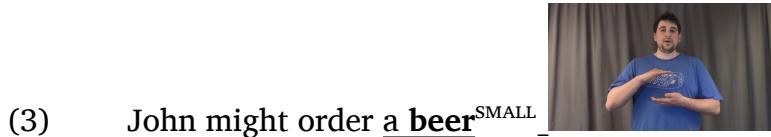
It has been further noted (Esipova 2018; the original observation about examples like (3) is due to Rob Pasternak (p.c.)) that co-speech gestures can in principle be interpreted as at-issue restrictive modifiers, in particular, under contrastive focus:

---

<sup>1</sup> In this paper I use the following notational conventions:

- In verbal expression<sup>GESTURE</sup> the gesture co-occurs with the verbal expression; the underlining loosely indicates the temporal alignment of the gesture, without making any syntactic claims.
- Gestures are sometimes illustrated by pictures after an underscore. The illustrations used throughout the paper are stills from the video stimuli used in the experiment.
- A **word** written in bold indicates prosodic contrastive focus marking (primarily, (L + )H\* pitch accent and lengthening on the stressed syllable).

<sup>2</sup> More precisely, that it would be roughly of the size indicated by the gesture, without necessarily making any commitments about whether that counts as large—for the sake of simplicity, I will be mostly ignoring this obvious caveat throughout the paper, though, and will be using such imprecise verbal equivalents of gestures as *small* and *large*.



→ If John orders a beer, it will be {small, large}.

≈ John might order a **small** beer or a **large** beer.

However, the actual acceptability status of examples like (3) has been unclear, even though different analyses of the semantics of co-speech gestures currently on the market make different predictions in this respect. The goal of the present study is to amend that.

### 1.1.1 Existing analyses of co-speech gestures

The exact semantic nature of inferences contributed by co-speech gestures, as in (1), is a matter of debate. Ebert (& Ebert) (2014; 2017) claims that co-speech gestures are Pottsonian (2005) supplements, akin to appositive relative clauses and nominal appositives. Throughout the paper I will refer to this analysis as *the supplemental analysis*.

Schlenker (2018) argues co-speech gestures trigger assertion-dependent conditional presuppositions (*cosuppositions*) of the form  $V \Rightarrow G$ , where  $V$  is the verbal expression the gesture adjoins to,  $G$  is the gesture's content, and  $\Rightarrow$  is generalized entailment. Those cosuppositions need to be satisfied in the local context of the complex word-gesture expression. I will refer to this analysis as *the cosuppositional analysis*.

The supplemental and cosuppositional analyses both assume that by default co-speech gestures make not-at-issue contributions, and in particular, that they preferably project from a variety of embedding environments, including from under *might*.<sup>3</sup> Thus, for (1) both analyses predict a projecting inference that if John orders a beer, it will be large (Ebert: (1) ≈ *John might order a beer, which (by the way) will be large*; Schlenker: in the local context of beer<sup>LARGE</sup> in (1), *beer* ⇒ *large*, which, given certain assumptions about how local contexts are computed, yields a presupposition roughly of the form 'If John brings a beer, it will be large').

<sup>3</sup> Though Tieu et al. (2017a; 2017b) found high endorsement rates even for inferences contributed by unambiguously at-issue modifiers under *might*. This is not a concern for the present study, since I don't look at inferential judgements.

The two analyses diverge in whether they allow for at-issue interpretations of gestures.

Schlenker (2018) claims that co-speech gestures can in principle have at-issue interpretations and attributes such interpretations to *local accommodation* of presuppositions, allowed for as a last resort in some standard theories of presupposition projection (Heim 1983; Schlenker 2009, a.o.).

One way of thinking about local accommodation is that the requirement, standardly imposed on presuppositions, that they be entailed by their local context, is lifted, and the presupposition is interpreted as a conjunct locally, as part of the at-issue content. An example of local accommodation for standard presupposition triggers is given below. Normally, *stopped V-ing* gives rise to the presupposition *used to V*, as is the case in (4). Conversely, *started V-ing* gives rise to the presupposition *used to not V*. Projecting both these presuppositions in (5) would result in a contradiction, so both presuppositions are locally accommodated under *maybe*.

- (4)     Maybe Zoe stopped smoking.  
          → Zoe used to smoke.
- (5)     A: Why is Zoe chewing on her pencil?  
       B: I don't know. Maybe she stopped smoking. Or maybe she  
          **started** smoking.  
          ↗ Zoe used to {smoke, not smoke}.  
          ≈ Maybe Zoe (used to smoke and stopped). Or maybe she  
          (used to not smoke and started).

Without going into technical details, under Schlenker's analysis, local accommodation of adnominal gestures that adjoin to NPs (i.e., constituents that denote predicates of individuals, like *beer* or *dog*, as opposed to DPs, which denote individuals or generalized quantifiers, like *a beer* or *John's dog*) would yield an ordinary restrictive modifier interpretation, i.e.,  $[\![\underline{I}_{NP} \underline{beer}]^{\text{LARGE}}]$  would be interpreted essentially as  $[\!\!large [\![\underline{I}_{NP} \underline{beer}]\!]]$ , which is exactly the reading claimed to be available in (3).

Local accommodation is typically taken to incur some cost<sup>4</sup>, the amount of which can vary across triggers (weak/soft vs. strong/hard triggers). One could thus envisage two versions of the cosuppositional analysis. Under the first version (which I believe is currently assumed by Schlenker himself),

---

<sup>4</sup> Throughout the paper I am making no commitments about the psycholinguistic nature of cost; I am using this term simply to refer to whatever results in lower acceptability. However, see Chemla & Bott 2013 for experimental data on response times as a measure of processing cost incurred by local accommodation.

gestural cosuppositions are triggered by default and can then be locally accommodated under some (possibly, minor) pressure, thus, incurring some cost. The amount of the cost will depend on the strength of co-speech gestures as triggers (Schlenker himself claims that co-speech gestures are weak triggers, thus, they can be locally accommodated relatively easily). I will refer to this version of the cosuppositional analysis as *the obligatory cosupposition analysis*.

Under the second version, gestural cosuppositions are only triggered given the right circumstances, thus, at-issue interpretations of co-speech gestures would be due to non-generation of presuppositions in the first place. Something along these lines is suggested in [Abusch 2010](#) for some very weak structural (as opposed to lexical) presuppositions, e.g., for existence presuppositions of *wh*-questions. This version would thus be compatible with a view that NP-level gestures are ordinary modifiers, which can in principle have restrictive or non-restrictive interpretations, depending on their content, the context, etc.—just like adjectives (see, e.g., [Leffel 2014](#) for a discussion of non-restrictive adjectives). The mechanism of the non-restrictive interpretations could still be very similar to Schlenker's cosuppositions, it's just that the restrictive interpretations wouldn't incur any special cost, since there will be no default triggering bias to overcome. Under this view, (1) (*John might order a beer<sup>LARGE</sup>*) is in fact roughly equivalent to (2) (*John might order a large beer*), and the modifier can equally easily have a not-at-issue, non-restrictive interpretation as well as an at-issue, restrictive one. I will refer to this version of the cosuppositional analysis as *the optional cosupposition analysis*.

Note that I remain agnostic about the specific source of the default triggering in the obligatory cosupposition analysis (as, to my knowledge, is Schlenker) and the nature of the cost of local accommodation. As pointed out by an anonymous reviewer, it can be due to some statistical learning: speakers observe that at-issue uses of co-speech gestures are rare and develop a bias against such interpretations, resulting in the defaultness of the projective interpretation. While, to my mind, this specific hypothesis would run into a sort of a chicken-and-egg dilemma (why are at-issue uses of gestures rare, if there is no inherent bias against them?), it would be compatible with my understanding of the obligatory cosupposition analysis. Crucially, the optional cosupposition analysis does not assume any bias whatsoever against at-issue interpretations of co-speech gestures, whether innate or developed via some statistical learning. In other words, within a stochastic setup, the difference between obligatory vs. optional cosupposition analyses wouldn't be about about the initial state of the grammar, but some sta-

ble final state thereof, which would have incorporated all the statistically learned biases.

Now, under the supplemental analysis, co-speech gestures shouldn't be able to have at-issue interpretations at all, since appositives typically don't, not even under pressure<sup>5</sup>:

- (6) #John might order a beer, which will be small, or he might order a beer, which will be large.  
 ≠ It might be that (John orders a beer and it is small), or it might be that (John orders a beer and it is large).

The example above is sharply infelicitous, because the two appositive relative clauses give rise to contradictory inferences, which can't be treated as at-issue conjuncts under the two instances of *might*, even though it would have been a perfectly sensible interpretation.

I summarize the predictions of the analyses above regarding at-issue interpretations of co-speech gestures in Table 1.

**Table 1:** Analyses of co-speech gestures.

analyses	supplemental	cosuppositional obligatory	optional
at-issue co-speech gestures	impossible	possible, with cost	possible, w/o cost

One final note regarding the analyses of co-speech gestures currently on the market is that in [Esipova 2018](#) I propose an essentially hybrid analysis whereby adnominal co-speech gestures are ordinary modifiers when they adjoin at the NP level, but they are appositive-like when they adjoin at the DP level. The NP vs. DP-level distinction is not relevant in the examples used in the present study, so I omit any detailed discussion of the predictions of this analysis. For the purposes of this study, its predictions are in line with either the obligatory or the optional cosupposition analysis, depending on further assumptions.

---

<sup>5</sup> For a more refined empirical argument to this effect see [Esipova 2018](#); I also discuss there why a counterpart of (6) with two restrictive relative clauses is not an instance of local accommodation of appositives, but a different structure altogether.

### 1.1.2 Bringing Contrastive Focus into the picture

In order to test the predictions above, we need a reliable way to force at-issue interpretations of co-speech gestures. I suggest doing so by making the gestures the only locus of contrast between two explicitly juxtaposed contrastive focus (CF) alternatives, as in the *beer<sup>SMALL</sup>* vs. *beer<sup>LARGE</sup>* example in (3).

The previous studies don't take into account the role of focus in how co-speech gestures are interpreted. Thus, Tieu et al. (2017a; 2017b) provide experimental data (truth-value judgements, picture selection tasks, inferential judgements) to support the claim that inferences contributed by co-speech gestures tend to project from a variety of embedded environments—significantly more so than contributions of control at-issue modifiers of the form *like this<sup>GESTURE</sup>* or alike:

- (7)     a. The boy will not use the stairs<sup>DOWN</sup>.  
        b. The boy will not use the stairs in this direction<sup>DOWN</sup>.

However, the endorsement rates of the purported gestural inferences in examples like (7a) were not always very high, which led Tieu et al. to the same conclusion as in Schlenker 2018, that co-speech gestures are weak presupposition triggers, which are susceptible to local accommodation under some relatively minor pressure (e.g., because the inferences contributed by the gestures might sometimes be too pragmatically odd to accommodate globally).

That said, the data from Tieu et al. don't really distinguish between the two versions of the cosuppositional analysis above; they show that co-speech gestures aren't *always* at-issue, but they don't tell us if co-speech gestures are significantly different from ordinary modifiers. The reason for that is that in (7b) the demonstrative *in this direction<sup>DOWN</sup>* is in focus, which presumably makes the PP obligatorily at-issue. More specifically, *The boy will not use the stairs in this direction<sup>DOWN</sup>* gives rise to a very natural alternative *The boy will use the stairs in this direction<sup>UP</sup>*, which the speaker arguably believes to be possible (if not true). My understanding is that, given certain assumptions about what alternatives the speaker believes to be possible, similar reasoning can apply to the other controls in Tieu et al.'s studies.

Compare this, for example, to how focus forces restrictive interpretations of adjectives (observation made for German in Umbach 2006 and extended to English in Leffel 2014):

- (8)     a. In Anna's garden there are colorful flowers.

- b. #In Anna's garden there are **colorful** flowers.

(Leffel 2014: (3.14))

The adjective in (8b) is forced to have a restrictive interpretation because of focus, which results in a somewhat odd sentence, because it suggests the existence of colorless flowers.

In (7a), however, focus is on *use the stairs<sup>DOWN</sup>*, and there is no reason why it would have to associate with the gesture rather than with the verbal expression. However, we can try to force focus to associate with co-speech gestures. In particular, this has to happen when we have contrastive focus markers on two (or more) complex word-gesture expressions, such that the gestures are the only locus of contrast between the two CF alternatives, as is the case in (3), repeated below.

- (3) John might order a **beer**<sup>SMALL</sup> or a **beer**<sup>LARGE</sup>.  
 ↗ If John orders a beer, it will be {small, large}.  
 ≈ John might order a **small** beer or a **large** beer.

Now, in (3) the gestures can only be interpreted as at-issue, regardless of CF. The two gestural inferences (predicted by both the supplemental and the cosuppositional analyses) are contradictory, so they can't both project, since that would mean requiring that the common ground entail a contradiction. Projecting only one of those inferences would make the alternative whose gestural inference doesn't project trivially false, which would make the whole utterance pragmatically odd, since it's odd to utter a disjunction one of whose disjuncts is known to be false.

In 2018, I argue that CF on complex word-gesture expressions forces at-issue interpretations of the gestures in other cases, too. However, in this paper I will restrict my attention to the more obvious case illustrated in (3).

## 1.2 Questions and predictions

The main goal of the present study is to investigate the acceptability of CF-forced at-issue interpretations of co-speech gestures. A secondary goal is to look at some factors that could affect that acceptability, namely, the content encoded by the gesture and the gesture's prosody.

### 1.2.1 Acceptability of at-issue co-speech gestures under CF (Contrast)

The target configuration to address the main question of this study is like in (3): contrasting two alternatives whose only locus of contrast is co-speech gestures. As a baseline for comparison I selected the configuration in which the verbal components of the CF-ed complex word-gesture expressions are contrastive but the gestures are not, as in (9), so that the at-issue interpretation of the gestures is not forced (although it might still be available).

- (9) John might order a beer<sup>SMALL</sup>   
 or a cocktail<sup>SMALL</sup> 

I will call the target and the baseline configuration *the Gestural Contrast condition* and *the Verbal Contrast condition*, respectively. The dimension along which the two differ will be referred to as *the Contrast factor*.

The non-contrastive gestures are added to the Verbal Contrast condition to partially compensate for the potential effect of CF markers co-occurring with non-contrastive material (verbal expressions) in the Gestural Contrast condition, although, of course, there might be an independent effect of non-contrastive gestures in the baseline configuration, since they are optional.

In Table 2 I supplement the previously adduced Table 1 with the specific predictions regarding the ratings of Gestural Contrast vs. Verbal Contrast examples. When formulating these predictions I make the following assumptions about how (im)possibility of certain interpretations translates to acceptability ratings:

- If an interpretation is impossible, i.e., not generated by the grammar (as is the case with at-issue interpretations of supplements), examples in which this interpretation is forced should receive low acceptability ratings.
- If an interpretation is possible, but comes at a cost (as is the case with local accommodation of presuppositions), the acceptability ratings of examples in which this interpretation is forced depend on the amount of the cost: the higher the cost, the lower the acceptability.
- If an interpretation is possible and comes at no cost (as is the case of non-generation of optional presuppositions), the acceptability rat-

ings of examples in which this interpretation is forced should be high (keeping in mind that participants in general rate examples with gestures relatively low, as shown in [Zlogar & Davidson 2018](#), etc.).

**Table 2:** Predictions of various analyses of co-speech gestures.

analyses	supplemental	cosuppositional	
		obligatory	optional
at-issue co-speech gestures	impossible	possible, with cost	possible, w/o cost
predictions for the Contrast factor	Gestural Contrast < Verbal Contrast; low absolute ratings for Gestural Contrast	Gestural Contrast < Verbal Contrast; absolute ratings for Gestural Contrast depend on the strength of co-speech gestures as triggers	Gestural Contrast = Verbal Contrast

### 1.2.2 Factors affecting acceptability of at-issue co-speech gestures under CF (Contrast/Content and Contrast/Emphasis)

An additional question I'm asking in this paper is what can affect the overall acceptability of CF-forced at-issue interpretations of co-speech gestures. With this question in mind I look at the interaction of Contrast with two further factors.

The first one is the type of content encoded by the gestures, in particular, whether said content is scalar or not. The *a priori* idea is that at-issue interpretations of scalar gestures under CF might be more acceptable, because once you evoke one alternative on a scale, the other alternatives might become particularly salient. An additional consideration is that sometimes one might want to be able to contrast two (or more) alternatives on a scale without making any absolute commitments (e.g., by assuming that a certain size counts as small or large), which might encourage using at-issue gestures instead of at-issue verbal modifiers.

To test this hypothesis I look at gestures encoding size (scalar content) and shape (non-scalar content). I will refer to the two conditions as *the Size*

*condition* and *the Shape condition*, respectively, and to the dimension along which the two differ as *the Content factor*.

For this factor the following hypotheses can be formulated, along with their predictions:

- Null hypothesis: no Contrast/Content interaction.
  - Claim: scalarity of gestural content has no effect on the acceptability of at-issue interpretations of co-speech gestures under CF.
  - Prediction: there is no interaction between Contrast and Content.
- Non-null hypothesis: Size Gestural Contrast > Shape Gestural Contrast.
  - Claim: scalarity of gestural content makes at-issue interpretations of co-speech gestures under CF more acceptable.
  - Prediction: in the Gestural Contrast condition Size examples enjoy higher acceptability than Shape examples.

The other factor I look at in this study is the prosodic properties of the gestures themselves. The idea behind this inquiry is that in order for co-speech gestures to have at-issue interpretations in the Gestural Contrast condition, prosodic CF markers have to associate with them rather than with the verbal material. That said, there might be a mismatch in how easily vocal prosodic markers can associate with vocal vs. non-vocal material. One could thus entertain the idea that putting more kinetic emphasis on a gesture might make the association of CF with it more acceptable.

To test this hypothesis I look at gestures produced with and without accelerated movement (*the Emphatic condition* and *the Non-Emphatic condition*, respectively). The dimension along which the two differ will be referred to as *the Emphasis factor*.

For this factor the following hypotheses can be formulated, along with their predictions:

- Null hypothesis: no Contrast/Emphasis interaction.
  - Claim: emphasis on a gesture has no effect on how easily CF can associate with it.
  - Prediction: there is no interaction between Contrast and Emphasis.
- Non-null hypothesis: Emphatic Gestural Contrast > Non-Emphatic Gestural Contrast.
  - Claim: emphasis on a gesture makes it easier for CF to associate with it.

- Prediction: in the Gestural Contrast condition Emphatic examples enjoy higher acceptability than Non-Emphatic examples.

## 2 Experiment

To test the hypotheses laid out above I conducted an acceptability judgement experiment whose design and results I report and discuss in this section.

### 2.1 Methods

#### 2.1.1 Participants

Participants were recruited via Amazon Mechanical Turk (MTurk), and were paid \$1.50 each for completing the study. Three participants were excluded because they reported not being native speakers of English. One more participant was excluded for failing all the attention checks. The remaining total number of participants was 104.

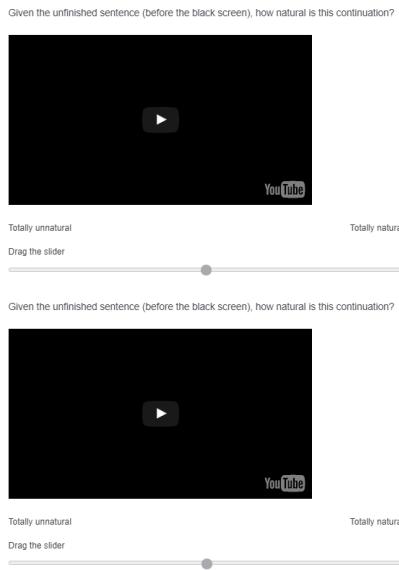
#### 2.1.2 Procedure

After accepting the MTurk task, participants were directed to an acceptability rating task hosted on Qualtrics. In each trial they watched two videos. Each of the two videos contained an unfinished sentence produced by a native speaker of English, which was the same across the two videos, followed by a continuation, which was different across the two videos. The unfinished sentence was separated from the continuations by a brief black screen.

Each video was accompanied by a slider scale whose left and right ends were labeled ‘Totally unnatural’ and ‘Totally natural’, respectively. Participants were instructed to rate the naturalness of each continuation by dragging the slider to the desired position on the scale. By default the slider was set to the middle position so as not to bias participants towards very low or very high ratings. While the scale seen by participants contained no numerical values, each position of the slider was mapped to a point on a 0–100-point scale.

Figure 1 shows the layout of a typical trial before the participants start playing the videos (the preview of a typical test item is the black screen

between the unfinished sentence and the continuation). See Appendix A for the specific instructions given to participants.



**Figure 1:** Layout of a typical trial (not to scale).

### 2.1.3 Materials

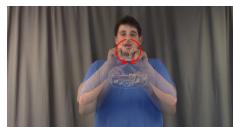
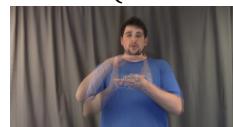
The design of the experiment is summarized in Table 3.

**Table 3:** Experiment design<sup>6</sup>.

<b>Content</b>	<b>Emphasis</b>	<b>Contrast</b>	
		<i>Gestural</i>	<i>Verbal</i>
<i>Shape</i>	<i>Emphatic</i>	4	4
	<i>Non-Emphatic</i>	4	4
<i>Size</i>	<i>Emphatic</i>	4	4
	<i>Non-Emphatic</i>	4	4

<sup>6</sup> Factors are bolded, conditions are italicized, the number in each cell indicates the number of test items.

For the experiment I constructed eight example pairs. Each example pair involved an unfinished sentence continued in two different ways, corresponding to the Gestural Contrast and the Verbal Contrast conditions (Contrast factor). There were four example pairs that contained Shape gestures and four that contained Size gestures (Content factor). A sample Shape example pair is given in (10) (a sample Size example pair was given before in (3) and (9)). Please find the full list of examples in Appendix B.

- (10) Kate only collects ashtrays<sup>SQUARE</sup> ...
- 
- a. ...she doesn't collect ashtrays<sup>ROUND</sup>.
- 
- (Gestural Contrast)
- b. ...she doesn't collect coasters<sup>SQUARE</sup>.
- 
- (Verbal Contrast)

A native speaker of English was then recorded producing the examples.

For each example two videos were recorded: with and without emphasis on the gestures (Emphasis factor). Since emphasis on trace gestures (in particular, the gesture ROUND) was found unnatural during the pilot stage, all the Shape items in the Emphatic condition only contained emphasis on the SQUARE gesture, which was uniformly the first gesture in each Shape example.

So, all in all, 32 videos were recorded: 8 Contrast-differing pairs (among which 4 with Size gestures, and 4 with Shape gestures), i.e., 16 sentences, and each sentence was recorded twice, with and without emphasis.

The resulting videos were then split into an unfinished sentence and a continuation and spliced back to make sure the unfinished sentence in each Gestural/Verbal Contrast pair of items was exactly the same. A one-second black screen was added between the unfinished sentence and the continuation in each video.

Additionally, the videos were edited so that the Emphatic and the Non-Emphatic versions of the same example contained the same audio track, to assure no interference from potential differences in vocal prosody in the

Emphasis factor.<sup>7</sup> The resulting stimuli were shown to several people who were familiar with the details of the experiment's design and several people who weren't, and no one noticed any lip-sync problems.

Please find all the videos used in the experiment, including the example videos from the instructions, here: <https://tinyurl.com/at-issue-gestures-stimuli>.

Let me add a quick note on non-manuals. While the speaker who was recorded was asked to try to keep his facial expressions and head movements consistent across different conditions (complete lack of those was assumed to be unnatural and hard to maintain without affecting other components of production), there was no further manipulation of the stimuli to assure said consistency. The role of eyebrow and head movements for CF marking has been discussed quite a lot in the literature, to name a few relevant studies: [Graf et al. 2002](#) and [Dohen et al. 2006](#) for production in English and French, respectively; [Krahmer et al. 2002](#), [Dohen & Loevenbruck 2009](#), and [House et al. 2001](#) for perception in Dutch, French, and Swedish, respectively. While the evidence regarding the importance of such non-manuals for perception of prominence in general and CF in particular is still inconclusive, it is in principle plausible that non-manuals could have interfered, in particular, with the Emphasis factor. That said, the eventual results for the Emphasis factor suggest that this is probably not a concern.

Each participant saw the same set of items. The items were organized into two fixed blocks, each containing eight item pairs, counterbalanced across the Content and the Emphasis factors. The order of presentation of the two blocks was randomly assigned to the participants. The order of presentation of the item pairs within each block was randomized for each participant. The order of presentation of the two items (Gestural Contrast vs. Verbal Contrast) was randomized for each trial.

Additionally, each block contained an extra trial that was an attention check (so, the total number of trials that each participant saw was 18). The videos in the attention check contained text instructions to drag the slider to the leftmost or the rightmost position on the scale. The attention check items looked exactly like test items; in particular, the previews of the videos were a black screen, which was also the case for most test items. Due to a

---

<sup>7</sup> The speaker producing the stimuli noted that it was often hard for him to produce non-emphatic gestures while putting vocal prosodic CF markers on the co-occurring verbal material. I looked at the pitch tracks in Praat ([Boersma & Weenink 2017](#)) and didn't notice any substantial differences in vocal prosody between the Emphatic and the Non-Emphatic versions, but no quantitative analysis has been done. The effect of gestural emphasis on production and perception of prominence is, of course, of huge interest, though.

technical glitch, this wasn't the case for all test items, but the participants still couldn't have known which items were attention check items based on the previews.

## 2.2 Results

All statistical tests and plots were done using R (R Core Team 2017). Here is a link to download a datasheet with the raw data: [http://esipova.net/files/esipova-at-issue\\_gestures-raw.csv](http://esipova.net/files/esipova-at-issue_gestures-raw.csv).

### 2.2.1 Results across all participants

In this subsection I report the results regarding the overall main effects of various factors and their interaction on acceptability ratings. Homogeneity of variance is not a concern for me, since my sample sizes for all conditions are equal (see, e.g., Cohen 2014, Chapters 10:B, 12:B), so, even though Levene's test returned a significant result for the Content factor ( $F(1, 3326) = 3.86, p = 0.0495$ ), we could still proceed with ordinary statistical tests.

The data exhibited a lot of variation across individual participants within the Contrast factor as well as across items (discussed in greater detail in the next section). For that reason I ran a linear mixed effects regression model with Contrast, Content, Emphasis, Contrast/Content interaction, and Contrast/Emphasis as fixed effects and Participant and Item as random effects (random intercepts and random slopes for Contrast). The statistics are summarized in Table 4; the  $p$  values reported were obtained using Satterthwaite's method.

Gestural Contrast examples ( $M = 39.10$ ) turned out to be significantly less acceptable than Verbal Contrast examples ( $M = 53.60$ ). Content and Emphasis did not have a significant effect. These results are visualized in Figure 2. Neither of the interactions turned out to be significant.

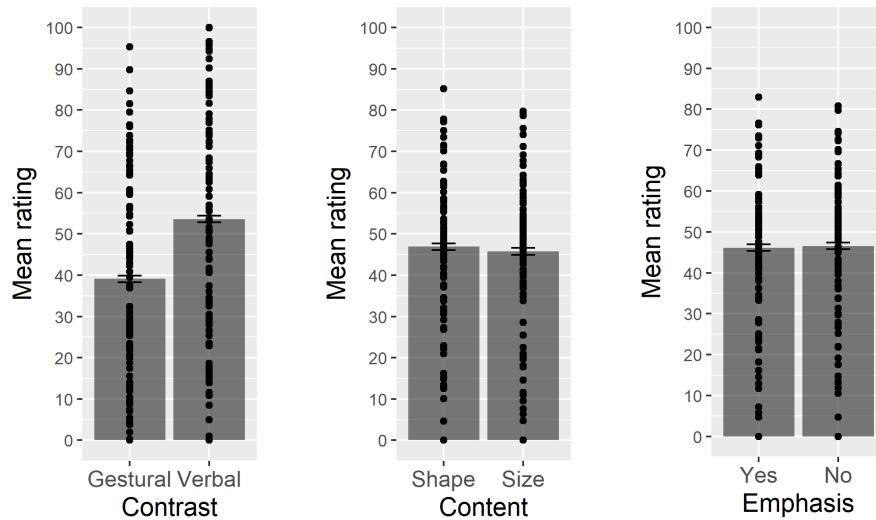
### 2.2.2 Variation across participants and items

While we could try to interpret the results for the Contrast factor from the previous section as it is, that would be misleading due to the high level of internally consistent variation across individual participants within this factor.

In fact, the interaction of Participant and Contrast is the best predictor in accounting for the variance in the data. A regression model with only

**Table 4:** Summary of the statistics.

		Beta (standardized)	t	p
Main effects	Contrast**	0.263	3.264	0.002
	Content	0.022	0.408	0.690
	Emphasis	0.007	0.134	0.896
Interactions	Contrast/Content	-0.069	-1.235	0.230
	Contrast/Emphasis	-0.043	0.966	

**Figure 2:** Mean ratings of Contrast, Content, and Emphasis. The bars show the mean ratings in the corresponding conditions across all participants, and the dots show individual mean ratings.

Contrast as a fixed effect—the only significant factor from the model in the previous section—accounts for only about 5% of the variance. A regression model with only Participant as a fixed effect is significantly different from an intercept only model, according to the likelihood ratio test, and accounts for about 22% of the variance. A model with both Contrast and Participant as fixed effects accounts for about 27% of the variance. Further adding the interaction of these two factors to this model as a fixed effect produces a significantly different model and boosts the amount of variance explained to about 65.5%.

Conversely, adding the interaction of Content and Participant as a fixed effect to a model that has these two factors as fixed effects does not pro-

duce a significantly different model. The same is true about Emphasis and Participant.

A model with Item as a fixed effect is also significantly different from an intercept only model and explains about 7% of the variance.

To summarize, participants vary a lot within the Contrast factor, but they don't vary within Content or Emphasis. In addition, there is a significant amount of variation across items, although it's not as drastic as individual variation.

These results are summarized in Table 5, where I report adjusted  $R^2$  for each model of interest and whether or not this model is significantly different from a model without the bolded effect based on the results of the likelihood ratio test.

**Table 5:** Fixed effects models showing the role of Participant and Item.

	Model	adj. $R^2$	p
1	Rating ~ <b>Participant</b>	0.219	<2.2e-16 ***
2	Rating ~ <b>Contrast</b>	0.049	<2.2e-16 ***
3	Rating ~ Participant + Contrast + <b>Participant*Contrast</b>	0.625	<2.2e-16 ***
4	Rating ~ Participant + Content + <b>Participant*Content</b>	0.209	0.998
5	Rating ~ Participant + Emphasis + <b>Participant*Emphasis</b>	0.202	1
6	Rating ~ Item	0.069	<2.2e-16 ***

The large amount of variance accounted for by the interaction between Participant and Contrast already suggests that the individual variation we observe for the Contrast factor isn't just noise, but instead, while individual judgement patterns vary a lot, participants are internally consistent in their judgement patterns.

Further quantitative evidence to that effect is the fact that the average inter-item correlation is 0.6 for Gestural Contrast examples and 0.65 for Verbal Contrast examples. For comparison, the average inter-item correlation across the whole data set is only 0.24 (unsurprisingly so, given that most participants rate Gestural and Verbal Contrast examples differently), and splitting the data along the Content or Emphasis dimensions doesn't boost this value at all.

To sum up, participants do vary a lot within the Contrast factor, but they do so in an internally consistent way.

## 2.3 Discussion

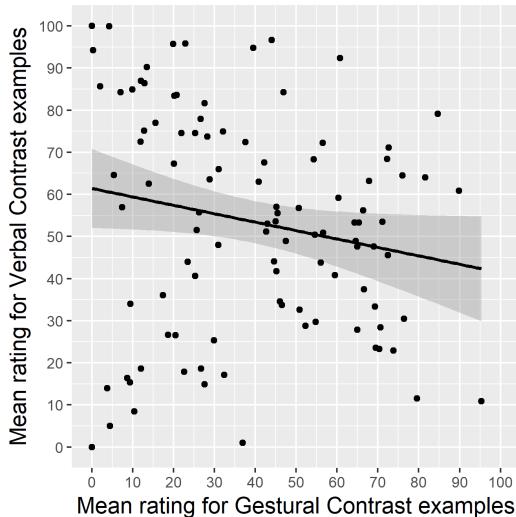
### 2.3.1 Contrast

The overall results for the main effect of the Contrast factor suggest that Gestural Contrast examples are less acceptable than Verbal Contrast examples. This would support the view that co-speech gestures by default contribute not-at-issue content, and making them at-issue is impossible or comes with a substantial cost, which is in line with the supplemental and obligatory co-supposition analyses, and contradicts the optional cosupposition analysis. The overall mean for the Gestural Contrast examples is quite low (39.1), but it's a little bit hard to interpret, since we don't have a baseline for ordinary presuppositions (including triggers that are typically considered weak and those that are typically considered strong) or supplements. Furthermore, the overall mean for the baseline Verbal Contrast examples is also quite low (53.6), which brings to mind the findings in Zlogar & Davidson 2018 that examples with co-speech gestures in general have lower acceptability ratings than their counterparts without co-speech gestures. In other words, the acceptability rating baseline for examples with co-speech gestures as such can be quite low to begin with.

All that said, this result for the Contrast factor shouldn't be interpreted straight-forwardly, since it is composed of highly variable, but internally consistent individual judgement patterns. Before we discuss the theoretical implications of this variation, let us first talk about its potential sources. To do so, let's take a look at what judgement patterns we observe in the data for the Contrast factor.

These patterns are visualized in Figure 3, where each dot represents an individual participant, with its position on the X axis being that participant's mean rating for Gestural Contrast examples and its position on the Y axis being their mean rating for Verbal Contrast examples. Thus, the dots in the top left sector of the plot represent participants who consistently rated Verbal Contrast examples higher than Gestural Contrast examples (they are the most numerous and drove the overall effect); the dots in the bottom right sector represent participants who consistently rated Gestural Contrast examples higher than Verbal Contrast examples, etc.

Note that a linear regression test for the two measures across all participants returned a borderline insignificant result:  $F(1, 102) = 3.83, p =$



**Figure 3:** Individual variation in acceptability of Gestural Contrast (X axis) vs. Verbal Contrast (Y axis) examples.

0.053, adjusted  $R^2 = 0.027$ . In other words, it's hard/impossible to predict a random individual's mean rating for Verbal Contrast examples based on their mean rating for Gestural Contrast examples (and vice versa), which suggests that speakers vary across these two dimensions independently.

One way to explain the variation at hand is to posit two independent dimensions of variation across individual *grammars*: (i) how high the cost of making co-speech gestures at-issue is for a given speaker, and (ii) how acceptable a given speaker finds optional non-contrastive material (in particular, gestures) under CF.<sup>8</sup> Variation along the first dimension will give us the speaker's position on the X axis of Figure 3, and variation along the second dimension will give us their position on the Y axis. Thus, these two dimensions of variation could be in principle enough to cover the entire variation space in Figure 3.

It is, however, also possible to attribute part of the variation observed to individual differences in *behavior* rather than grammar, in particular, how willing a given speaker is to ignore the contribution of co-speech gestures altogether when performing an acceptability judgement task. Tieu et al. (2017a; 2017b), for example, claim that the possibility that some participants were ignoring the gestures could explain some of their data. However,

<sup>8</sup> There is, of course, a third potential dimension of variation, which is how acceptable a given speaker finds examples with co-speech gestures to begin with, but the present study doesn't offer any access to this information.

Tieu et al. don't discuss individual variation in their data—it would be interesting to look at the individual patterns of judgements in their data to see if indeed that was a possibility.

Behavioral variation alone wouldn't explain all the variation observed in this study, though. To see that, let us imagine what the variation space from Figure 3 would look like if there was no variation along dimensions (i) and (ii) suggested above, and the only locus of variation was when, if at all, a given participant chooses to ignore the gestures.

The judgements of the participants who never ignore the gestures—let's call them Group A—should depend on the cost of making co-speech gestures at-issue (X axis) and on the cost of having non-contrastive optional material under CF (Y axis). If there is no variation along these two dimensions, all such participants should cluster in the same area of the variation space.

Ignoring the gestures in the Gestural Contrast condition should result in entirely deviant sentences with completely non-contrastive CF alternatives (e.g., *John might order a beer or a beer*). Ignoring the gestures in the Verbal Contrast condition should result in entirely acceptable sentences (from the pure grammaticality point of view), with perfectly contrastive CF alternatives and without non-contrastive gestures (e.g., *John might order a beer or a cocktail*).

Thus, participants who choose to ignore the gestures altogether across all conditions—let's call them Group B—should cluster in the very top left corner of the variation space from Figure 3.

It is also possible that some participants adopt a differential behavior: they ignore the gestures in the Verbal Contrast condition, when ignoring the gestures makes the sentence completely acceptable, but they don't ignore the gestures in the Gestural Contrast condition, when ignoring the gestures makes the sentence completely unacceptable. Such participants would then cluster near the top edge of the variation space, but their position on the X axis should align with that of Group A.

A reverse pattern (when a participant ignores the gestures in the Gestural Contrast condition but not in the Verbal Contrast one) is technically possible, but unlikely, assuming that participants in general are more likely to look for ways to make an example more acceptable than more degraded. That said, if some people were to adopt this “antagonistic” pattern—let's call them Group D—they would cluster near the left edge of the variation space, but their position along the Y axis should align with that of Group A.

To sum up, if there is no variation along either of the two grammatical dimensions, we expect at most four clusters in the variation space: Group A in some position in the variation space, Group B in the top left corner,

Group C near the top edge and aligned with Group A along the X axis, and Group D near the left edge and aligned with Group A along the Y axis. In other words, we expect to see at most four clusters forming a square, or some subset of its vertices. If there is variation along only one of the two grammatical dimensions, we expect two columns or two rows of dots. It is easy to see that Figure 3 cannot possibly satisfy either of these scenarios.

Thus, the supposition that some participants simply ignored the gestures can't explain all the variation observed for the Contrast factor. Therefore, I believe it is reasonable to assume some amount of grammatical variation along both dimensions (i) and (ii) suggested above. Of course, it is still entirely possible that some other behavioral variation that I haven't considered is at play.

The reasoning above, of course, doesn't prove that no participants ignored the gestures. Unfortunately, the current data do not really allow us to see if/when a given participant ignored the gestures, except for (quite few) individual participants who chose to leave informative comments at the end of the study (the comments were optional), from which it was clear they were not ignoring the gestures. One way to get at the relevant information in potential follow-up studies would be to ask people directly if/when they ignored the gestures and make that question obligatory. A more indirect way would be to ask people inferential questions about the contribution of the gestures, but that will be informative only in a subset of cases, in particular, only in the Verbal Contrast condition and only when a given participant reports projection of the gestural inferences. In the Gestural Contrast condition the gestural inferences aren't supposed to project, and if someone doesn't get projection in the Verbal Contrast condition, it is possible that they don't ignore the gestures but treat them as at-issue for independent reasons.

Setting that issue aside, assuming that we are at least to some extent dealing with bona fide grammatical variation, let's look at what further theoretical insights regarding the semantics of co-speech gestures the observed pattern of variation along the X axis can offer.

The fact that the right edge of the variation space in Figure 3 is very scarcely populated suggests that not many people, if any, treat co-speech gestures as ordinary modifiers that are freely ambiguous between non-restrictive and restrictive interpretations, contra the optional presupposition analysis. In other words, it looks like most people's grammars do have a bias against at-issue interpretations of co-speech gestures.

Such bias is present both in the supplemental and obligatory cosuppositional analyses. However, the supplemental analysis predicts categorical

unacceptability of at-issue interpretations of co-speech gestures (at least under the assumption that all supplements behave uniformly with respect to the availability of at-issue interpretations), in which case we would expect the speakers to cluster along the left edge of the variation space from Figure 3, which is clearly not the case. I take this to suggest that this version of the supplemental analysis is not tenable. More generally, no analysis with a fixed cost of at-issue interpretations of co-speech gestures would be able to capture the empirical picture observed in this study.

As for the obligatory cosupposition analysis, the assumption that presupposition triggers differ in strength, i.e., in how easily they allow local accommodation, is already ubiquitous in the literature, thus, the nature of the cost of local accommodation is already gradient. However, one would also need to adopt the view that speakers can vary in the strength they assign to co-speech gestures as presupposition triggers. Alternatively, one could posit that speakers vary in how ready they are to accept local accommodation in the first place, even for what's typically considered weak triggers (this option doesn't exclude the one above, of course).

To distinguish between the two possibilities, it would be good to conduct further studies looking at if there is any correlation in how a given speaker treats ordinary presupposition triggers (of various alleged strength) and co-speech gestures with respect to local accommodation. More generally, looking at the amount of inter-speaker variation (both grammatical and behavioral) regarding the cost of interpreting some content that is typically not-at-issue as at-issue for gestures vs. other types of not-at-issue content can be potentially very illuminating in view of the discussion on how linguistically integrated gestures are, the intuitive idea being that the less grammaticalized a certain phenomenon is the more variation one would expect in how individual speakers treat it.

I should note that there is also a possibility that it is not just the cost of making co-speech gestures at-issue that makes the Gestural Contrast examples in the present study degraded for many people. As mentioned before, having CF markers co-occur with non-contrastive material, even if it's not optional, might be inherently marked. I can envisage two potential hypotheses in this respect: (i) a phonological one: having CF markers co-occur with two identical phonological strings is marked, and (ii) a semantic one: having CF markers co-occur with two semantically identical chunks, even if those markers don't associate with those chunks, is marked. The two, of course, don't contradict each other and can in principle have a cumulative effect.

Hypothesis (i) could be tested on its own by measuring the acceptability of phonologically identical pronouns and indexicals co-occurring with contrastive pointing under CF, as in (11), where the semantic content of the two pronouns/indexicals is arguably already contrastive (and the pointing gestures are there just to help identify the referents), but their phonological form is identical.

- (11) a. I like him<sup>POINT-A</sup>, but I don't like him<sup>POINT-B</sup>.  
 b. I like this<sup>POINT-A</sup> book, but I don't like this<sup>POINT-B</sup> book.

It is unclear to me how to independently test hypothesis (ii), though. For that we would need two expressions that have contrastive phonology but identical semantics (and have them co-occur with contrastive gestures), which, even if we believe in full synonyms as such, will most likely inevitably lead to metalinguistic interpretations.

In this respect it would be good to do follow-up studies that would look at whether speakers who are likely to obtain at-issue interpretations of co-speech gestures without CF (as in Tieu et al.'s results) are also more accepting of CF-forced at-issue interpretations.

One final note in this subsection is that in the present study no demographic data were collected (other than on the languages that a given participant speaks), so there is no way to assess, for example, the effect of age on how readily a given participant accepts at-issue interpretations of co-speech gestures under CF. In follow-up studies it would be best to collect such data to see if there are any sociolinguistic tendencies of interest.

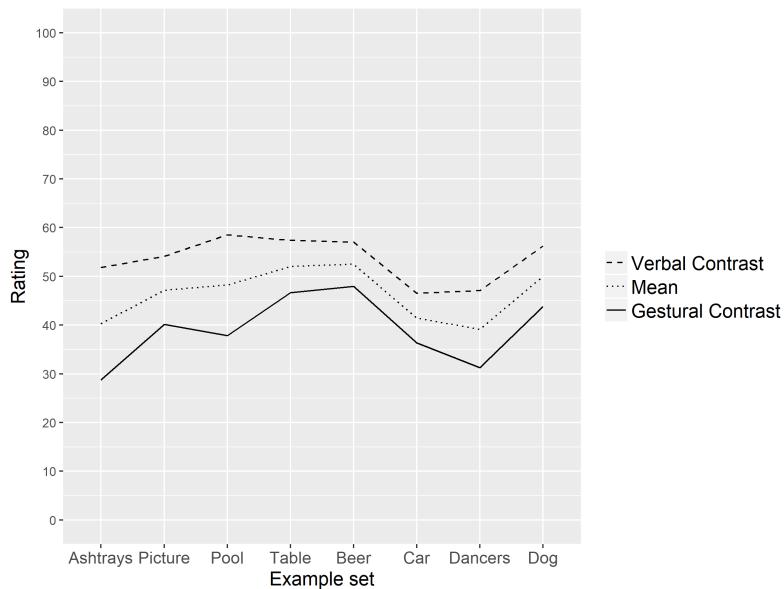
### 2.3.2 Content/Contrast

The results for the Content/Contrast interaction don't support the non-null hypothesis that scalarity of content makes at-issue interpretations of co-speech gestures under CF more acceptable.

That said, in the present study I only looked at two—rather broad—types of content, size and shape. It is in principle possible that the type of content does play a role in how acceptable it is to use a gesture as an at-issue modifier, but in a more idiosyncratic way. That would be in line with the significant amount of variation we observed across items.

Following the practice in Zlogar & Davidson 2018, who also observed a lot of variation across items, below I report the mean ratings within different example sets for methodological reasons, since these data could be helpful for subsequent experiments on gestures.

Figure 4 shows the mean ratings across all participants for the sets of examples used in the experiment (see the list of examples in Appendix B). The points on the dotted (middle) line, labeled ‘Mean’, represent the mean ratings for the example sets across all conditions. The dots on the solid (lower) line, labeled ‘Gestural contrast’, represent the mean ratings for the Gestural Contrast examples within those example sets. The dots on the dashed (higher) line, labeled ‘Verbal Contrast’, represent the mean ratings for the Verbal Contrast examples within those example sets. The first four sets on the X axis (Ashtrays, Picture, Pool, Table) contain Shape examples, and the other four sets (Beer, Car, Dancers, Dog) contain Size examples.



**Figure 4:** Variation across sets of examples.

The plot thus illustrates two aspects of variation. First, all individual examples vary in absolute acceptability: within the Gestural Contrast condition, the value range is 28.77–47.94, and within the Verbal Contrast condition the range is 46.52–58.53.

Second, while for all example sets the mean rating for the Verbal Contrast example is higher than the mean rating for the Gestural Contrast example, there is variation in the size of the gap between the two. The value range for this gap is 9.13–23.07.

### 2.3.3 *Emphasis/Contrast*

The null result for the Emphasis/Contrast interaction doesn't allow us to reject the null hypothesis that kinetic emphasis on co-speech gestures has no effect on the acceptability of at-issue interpretations of those gestures under CF (or on the acceptability of non-contrastive gestures under CF for that matter).

One obvious possibility is that speakers aren't attuned to such subtle differences in the first place. Another possibility is that speakers are attuned to such differences, but they play no role in the acceptability of at-issue interpretations of co-speech gestures under CF. Finally, it is also possible that speakers are attuned to such differences, and they can in principle affect the acceptability of at-issue interpretations of co-speech gestures under CF, but the participants of the present study chose to ignore the contribution of the gestural emphasis for the purposes of the task at hand.

The data obtained in this experiment don't really allow us to distinguish among the possibilities above, although four speakers left comments implying that they thought that the videos in the two blocks were the same, which suggests that at least those participants weren't consciously aware of the difference.

This is in line with the reaction I obtained during the preparation stage from five native speakers of English (all linguists), who were shown the emphatic and non-emphatic versions of the same example and were asked directly what the difference between the two was. Even though all those speakers were to a varied extent familiar with the goals of the experiment, they couldn't immediately tell what the difference was. However, some of them suggested that there were differences in vocal prosody between the two. For example, some of the comments were that the word co-occurring with an emphatic gesture had "a higher pitch" or "a more emphatic intonation" (once again, the audio track was the same in the two versions). It is possible that those speakers did subconsciously notice the difference between the emphatic and non-emphatic gestures and perceived the strings with the emphatic gestures as overall more prominent, but then mis-attributed the higher prominence to something in vocal prosody (pitch, loudness, etc.). It would be independently interesting to investigate this effect further.

One final note in this respect: Amir Anvari (p.c.) pointed out to me that while kinetic emphasis might not play a role in how acceptable a given gesture is as an at-issue modifier, other ways of making a gesture more salient, such as producing it closer to one's face rather than in the neutral

gestural space, might. This supposition is worth investigating in follow-up studies.

### 3 Conclusion

In this study I used an acceptability judgement task to investigate the acceptability of at-issue interpretations of co-speech gestures forced by CF.

The overall results show that sentences in which at-issue interpretations of co-speech gestures are forced to make CF felicitous are degraded, in particular, when compared to controls in which at-issue interpretations of co-speech gestures are not forced. These findings support the view that co-speech gestures by default make not-at-issue contributions, and making them at-issue is costly, which is broadly compatible with several existing analyses of co-speech gestures: the supplemental analysis (Ebert & Ebert 2014; Ebert 2017), the cosuppositional analysis (Schlenker 2018) with obligatory triggering of gestural cosuppositions and possibility of costly local accommodation under pressure, and some versions of the hybrid analysis in Esipova 2018.

However, I also observed a high amount of variation in individual judgement patterns. Looking at said variation proved illuminating, since it provided evidence that co-speech gestures cannot be uniformly treated as supplements (under the assumption that supplements uniformly don't allow at-issue interpretations)—a conclusion that can be overlooked if one only looks at the overall effects.

Furthermore, the variation data can be used to argue against any analysis in which the cost of making co-speech gestures at-issue is fixed across speakers. This raises a more general question about how much speakers can vary in how readily they accept at-issue interpretations of different types of typically not-at-issue content and whether the amount of such variation for a given type of not-at-issue content depends on how linguistically integrated it is.

I have also additionally looked at what factors can affect the acceptability of CF-forced at-issue interpretations of co-speech gestures.

The results regarding the type of content encoded by the gesture suggest that there is no difference between size and shape gestures when it comes to the acceptability of at-issue interpretations. Follow-up studies could focus on more fine-grained content type distinctions.

It was further found that emphatic production of co-speech gestures (in particular, producing a gesture with accelerated movement, as opposed to

producing it with no movement) does not affect the acceptability of at-issue interpretations of those gestures. Follow-up research could focus on distinguishing among different potential reasons for this lack of effect or on whether other prosodic factors can affect the acceptability of such at-issue interpretations.

## Abbreviations

CF = contrastive focus

## Ethics and consent

This study was approved and deemed exempt from a full review by the Institutional Review Board at New York University (IRB#: IRB-FY2017-707). For any further questions, you may contact the University Committee on Activities Involving Human Subjects, New York University at 212-998-4808 or [ask.humansubjects@nyu.edu](mailto:ask.humansubjects@nyu.edu).

## Acknowledgements

I am very grateful the committee of my second qualifying paper, which served as a basis for this paper, for feedback and support at each stage of this project: Ailís Cournane (Chair), Kathryn Davidson, and Philippe Schlenker. Special thanks to Rob Pasternak for professional and patient help with the stimuli for this study, as well as for many fruitful conversations. For discussions, bits of advice, and/or other help at different stages of this project I'd also like to thank Lucas Champollion, Edward Flemming, Laurel MacKenzie, Jon Rawski, Adina Williams, everyone who participated in piloting the experiment, and the audience at the 2018 Annual Meeting of the LSA 2018. Finally, many thanks to the editorial team of 'Glossa', and especially Jo-han Rooryck, and three anonymous reviewers for helpful and constructive feedback.

## Competing interests

The author has no competing interests to declare.

## References

- Abusch, Dorit. 2010. Presupposition triggering from alternatives. *Journal of Semantics* 27(1). 37–80. <http://dx.doi.org/10.1093/jos/ffp009>.
- Boersma, Paul & David Weenink. 2017. Praat: doing phonetics by computer (version 6.0.29). <http://www.praat.org>.
- Chemla, Emmanuel & Lewis Bott. 2013. Processing presuppositions: Dynamic semantics vs pragmatic enrichment. *Language and Cognitive Processes* 28(3). 241–260. <http://dx.doi.org/10.1080/01690965.2011.615221>.
- Cohen, Barry H. 2014. *Explaining psychological statistics, 4th edition*. Wiley.
- Dohen, Marion & Hélène Loevenbruck. 2009. Interaction of audition and vision for the perception of prosodic contrastive focus. *Language and speech* 52(2-3). 177–206.
- Dohen, Marion, Hélène Loevenbruck & Hill Harold. 2006. Visual correlates of prosodic contrastive focus in french: description and inter-speaker variability. In *Speech Prosody*. 221–224.
- Ebert, Cornelia. 2017. Co-speech vs. post-speech gestures. Talk given at *Language and cognition workshop in memory of Peter Bosch*, Cognitive Science Department, University of Osnabrück, February 2017.
- Ebert, Cornelia & Christian Ebert. 2014. Gestures, demonstratives, and the attributive/referential distinction. Talk given at *SPE 7*, Berlin, June 28. <http://semanticsarchive.net/Archive/GJjYzkwN/EbertEbert-SPE-2014-slides.pdf>.
- Esipova, Maria. 2018. Focus on what's not at issue: gestures, presuppositions, supplements under contrastive focus. In *Proceedings of Sinn und Bedeutung 22*.
- Graf, Hans Peter, Eric Cosatto, Volker Strom & Fu Jie Huang. 2002. Visual prosody: Facial movements accompanying speech. In *Automatic Face and Gesture Recognition, 2002, proceedings of the fifth ieee international conference*. 396–401.
- Heim, Irene. 1983. On the projection problem for presuppositions. In Michael Barlow, Daniel Flickinger & Michael Wescoat (eds.), *Second West Coast Conference on Formal Linguistics*. 114–125.
- House, David, Jonas Beskow & Björn Granström. 2001. Timing and interaction of visual cues for prominence in audiovisual speech perception. In *Proceedings of the European Conferences on Speech Communication and Technology (Eurospeech)*. 387–390.
- Krahmer, Emiel, Zsófia Ruttkay, Marc Swerts & Wieger Wesselink. 2002. Pitch, eyebrows and the perception of focus. In *Speech Prosody*. 443–

- 446.
- Leffel, Timothy. 2014. *The semantics of modification: Adjectives, nouns, and order*: NYU dissertation. [http://lefft.xyz/pdf/leffel\\_2014\\_nyu\\_diss.pdf](http://lefft.xyz/pdf/leffel_2014_nyu_diss.pdf).
- Potts, Christopher. 2005. *The logic of conventional implicatures*. Oxford University Press.
- R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. <https://www.R-project.org>.
- Schlenker, Philippe. 2009. Local contexts. *Semantics and Pragmatics* 2(3). 1–78. <http://dx.doi.org/10.3765/sp.2.3>.
- Schlenker, Philippe. 2018. Gesture projection and cosuppositions. *Linguistics and Philosophy* <http://dx.doi.org/10.1007/s10988-017-9225-8>.
- Tieu, Lyn, Robert Pasternak, Philippe Schlenker & Emmanuel Chemla. 2017a. Co-speech gesture projection: Evidence from inferential judgments. Ms. [http://semanticsarchive.net/Archive/jBiMmUwM/TieuPasternakSchlenkerChemla\\_GestureInferences.pdf](http://semanticsarchive.net/Archive/jBiMmUwM/TieuPasternakSchlenkerChemla_GestureInferences.pdf).
- Tieu, Lyn, Robert Pasternak, Philippe Schlenker & Emmanuel Chemla. 2017b. Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa* 2(1). 102. <http://dx.doi.org/10.5334/gjgl.334>.
- Umbach, Carla. 2006. Non-restrictive modification and backgrounding. In *Proceedings of the Ninth Symposium on Logic and Language*. 152–159. Budapest: Hungarian Academy of Sciences.
- Zlogar, Christina & Kathryn Davidson. 2018. Effects of linguistic context on the acceptability of co-speech gestures. *Glossa: a journal of general linguistics* 3(1). 73. <http://dx.doi.org/10.5334/gjgl.438/>.

## Appendix A. Instructions

Here I adduce the instructions that the MTurk participants saw before completing the study (some of the aspects of the formatting, such as, for example, the fonts, were not preserved).

### Instructions

To complete this study you will need a computer with Internet connection and an up-to-date version of Adobe Flash Player, as well as headphones or speakers. Please make sure your screen is big enough to watch videos.

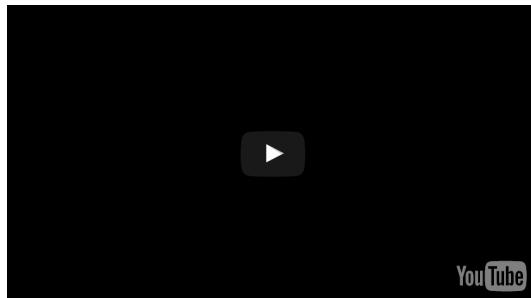
This study has 2 blocks of 9 trials each (i.e., 18 trials altogether), and should take about 12–15 minutes to complete. In each trial you will watch an unfinished sentence in English containing a gesture followed by two different continuations, each of them also containing a gesture. The unfinished sentence will be separated from each of the continuations by a brief black screen.

Your task will be to assess how natural each of the two continuations to the unfinished sentence is by dragging a slider on a continuous scale from ‘Totally unnatural’ to ‘Totally natural’. (If you want to leave the slider where it is, i.e., in the middle of the scale, you will need to at least click on it, otherwise you won’t be allowed to proceed.) **Please pay attention to the gestures in the videos when assessing the continuations.**

Please note that there are no right or wrong answers here. When assessing the naturalness of the continuations you are expected to rely on your own linguistic intuitions rather than some prescriptive rules of “good English”.

Here is an example (your judgements might be different from those provided in the example):

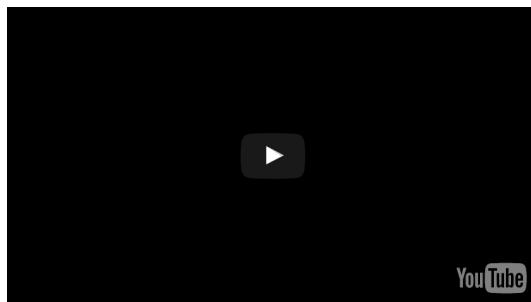
Watch an unfinished sentence followed by continuation A:



For most speakers of English this is a coherent, natural continuation to this unfinished sentence. If you are among those speakers, you might want to judge it very high on the naturalness scale, for example:



Now watch the same unfinished sentence followed by continuation B:



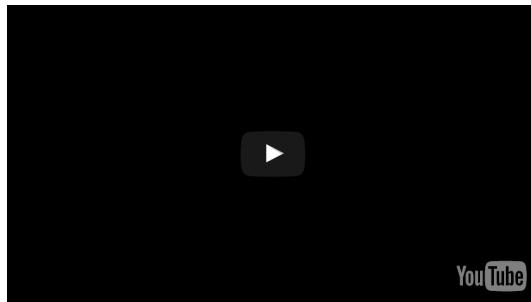
Continuation B could've been OK on its own, as a standalone utterance, but as a continuation to this unfinished sentence it is unnatural for most speakers of English, because the speaker in the video uses the same gesture for 'small' in the continuation that he used for 'large' in the unfinished sentence. If you, too, find it unnatural, you might want to judge it quite low on the naturalness scale, for example:

Totally unnatural                      Totally natural

Drag the slider

A horizontal slider control with a central circular handle. The text "Totally unnatural" is at the left end and "Totally natural" is at the right end. Below the slider is the instruction "Drag the slider".

Finally, watch the same unfinished sentence followed by continuation C:



This continuation is also weird for many speakers of English. In the unfinished sentence the speaker puts emphasis on 'large', suggesting that the continuation is likely to be about the size of the fish Mike did catch. The continuation, however, is about Alice. It is perhaps not unfathomable that the speaker might eventually decide to talk about the fish Alice caught, but the unfinished sentence still remains in some sense unfinished. Some speak-

ers might want to rate this sentence as marginal on the naturalness scale, but closer to the low end (once again, your judgements might be different), for example:



**In order to get paid, please make sure that you complete all 18 trials.**

## Appendix B. List of examples

Here I list the sentences used in the video stimuli and in the example videos for the instructions. No illustrations are provided for the gestures. Please find all the videos used in the experiment, including the example videos from the instructions, here: <https://tinyurl.com/at-issue-gestures-stimuli>.

### Example videos from the instructions

- (12) Mike didn't catch a large<sup>LARGE</sup> fish...
- a. ...he caught a small<sup>SMALL</sup> fish.
  - b. ...he caught a small<sup>LARGE</sup> fish.
  - c. ...Alice caught a large<sup>LARGE</sup> fish.

### Video stimuli

In each pair below (a) is a Gestural Contrast example, and (b) is a Verbal Contrast one.

#### *Shape gestures*

- (13) Kate only collects ashtrays<sup>SQUARE</sup>...
- a. ...she doesn't collect ashtrays<sup>ROUND</sup>
  - b. ...she doesn't collect coasters<sup>SQUARE</sup>.
- (14) On this wall Bill might hang a picture<sup>SQUARE</sup>...
- a. ...or a picture<sup>ROUND</sup>.

- b. ...or a clock<sup>SQUARE</sup>.
- (15) Anna's unlikely to build a pool<sup>SQUARE</sup> ...
- ...but she might build a pool<sup>ROUND</sup>.
  - ...but she might build a deck<sup>SQUARE</sup>.
- (16) In front of the couch Tim might put a table<sup>SQUARE</sup> ...
- ...or a table<sup>ROUND</sup>.
  - ...or an ottoman<sup>SQUARE</sup>.

*Size gestures*

- (17) John might order a beer<sup>SMALL</sup> ...
- ...or a beer<sup>LARGE</sup>.
  - ...or a cocktail<sup>SMALL</sup>.
- (18) Emma's unlikely to play with a car<sup>SMALL</sup> ...
- ...but she might play with a car<sup>LARGE</sup>.
  - ...but she might play with a helicopter<sup>SMALL</sup>.
- (19) They only hire dancers<sup>TALL</sup> ...
- ...they don't hire dancers<sup>SHORT</sup>.
  - ...they don't hire athletes<sup>TALL</sup>.
- (20) Dan might buy a dog<sup>SMALL</sup> ...
- ...but he's unlikely to buy a dog<sup>LARGE</sup>.
  - ...but he's unlikely to buy a cat<sup>SMALL</sup>.