RESEARCH

Effects of linguistic context on the acceptability of co-speech gestures

Christina Zlogar and Kathryn Davidson
Department of Linguistics, Harvard University, Boylston Hall, 3rd floor, Cambridge, MA 02138, US
Corresponding author: Christina Zlogar (czlogar@fas.harvard.edu)

We ask whether iconic co-speech gestures are judged as more natural by naive participants when their content is entailed by a preceding context, or repeated in the same utterance, or when they contribute new information (i.e., are nontrivial). Our results show, first, that the acceptability of co-speech gestures is not affected by whether they are entailed by a preceding context (they are not "hard presupposition triggers"). In contrast, our second finding is that gestures are affected by content in the same utterance, such that they are judged more acceptable when they reinforce information already present in the proposition. In addition, gestures varied widely in how acceptable they were judged even in the same information contexts, which we take to be an indication that we are just scratching the surface of a more general question about the felicity of co-speech gestures, of which the current study provides a first experimental foundation.

1 Introduction

One of the most productive developments of the last several decades of formal semantics/pragmatics research has been increasingly sophisticated models of different kinds of meaning that linguistic content can contribute, whether that content is at-issue or backgrounded, entailed or implicated, or updating or proposing updates to a shared common ground between interlocutors, among other distinctions. A classic three-way divide between entailments, implicatures, and presuppositions (see Chierchia & McConnell-Ginet 2000: Chapter 1 for a brief introduction) has been expanded in a number of ways through more complex distinctions. For example, although presuppositional elements like *too*, *again,* etc., often both require particular discourse conditions to be uttered and can "project" through other logical operators so as not to be affected by them, these two properties attributed classically to presuppositions can sometimes be dissociated, which means that attempts to model the contributions of these and other aspects of meaning should be careful to test each property separately.

Recently, this type of formal semantic/pragmatic analysis has been extended beyond language in a narrow sense to co-speech gesture (Ebert & Ebert 2014; 2016; Tieu et al. 2017; Esipova 2018; Schlenker 2018). On the face of it, this is a natural extension, given that co-speech gestures are known to be prosodically integrated with spoken language and have been argued to contribute to a unified semantic content together with speech/sign (McNeill 1992; Kendon 2004; Goldin-Meadow & Brentari 2017). However, there is some reason for caution about assuming that applications of the same linguistic tests for levels of semantic/pragmatic contribution can be applied wholesale to gestural content. One reason might be that speakers of a language like English may be comfortable making

metalinguistic judgments about speech, but less so about gesture, which is almost never a target of explicit instruction and so less frequently considered metalinguistically. Another reason is that gestural content may more often be interpreted in an analog way, potentially leading to more gradient grammaticality and/or truth value judgments. Finally, if there is some agreement among formal analyses of co-speech gestures, it is that they are frequently not-at-issue (Ebert & Ebert 2014; 2016; Schlenker 2018), and not-at-issue content can vary significantly across phenomena and across languages in several respects, including, as we mentioned above, how it projects through various logical operators and whether it imposes any restrictions on the previous discourse. In this paper we focus on this last aspect, by experimentally testing the sensitivity of co-speech gestures to linguistic context, construed broadly as both discourse context and the local context of the simultaneous speech stream. Our hope is that this will contribute an important foundational piece as part of a larger discussion regarding the semantic, pragmatic, and information structural properties of co-speech gestures, and the practical issues involved in creating appropriate contexts in which to test them via fieldwork or experimentation.

1.1 Background: Co-speech gestures

The kind of *co-speech gestures* that have been the focus of recent semantic/pragmatic work and that we will be focusing on in this paper are gestures made simultaneously with speech/sign that enrich utterance meaning by iconically depicting an aspect of the situation or event (Kendon 2004; Goldin-Meadow & Brentari 2017); see (1–6) for examples. In these and future examples, gestures will be written in all-capital letters, and we will indicate the spoken words that align with the gesture by placing them in square brackets.

(1)   McNeill (1992: 79)
        …and he [bends it way back]_PULL-BACK

(2)   Goldin-Meadow & Brentari (2017: 35)
        I [ran up the stairs]_SPIRAL-UP.

(3)   Ebert & Ebert (2016)
        One child managed to cut out [a geometrical form]_TRIANGLE.

(4)   Schlenker (2018: 2)
        John [helped]_UP his son.

(5)   Tieu et al. (2017: 3)
        I brought [a bottle]_LARGE to the talk.

(6)   She [scored]_SHOOT the winning point!

All of the co-speech gestures in (1–6) convey some information which the literal speech stream does not. For example, the co-speech gesture PULL-BACK in (1) involves the speaker making as if to grip and pull back something flexible which is fixed at the base, such as a sapling. The iconic information communicated by the gesture but not the spoken words is that the object being pulled back is fixed at the base. Similarly, the SPIRAL-UP gesture in (2) alone conveys the fact that the staircase in question is a spiral staircase, and the gesture TRIANGLE in (3) entails that the geometric shape in question is a triangle. UP in (4) implies that the helping was done by lifting, and LARGE in (5) implies that the bottle was large. Finally, the SHOOT gesture in (6) indicates that the scoring was done by shooting a basketball.

Kendon (2004: 7) defines a gesture as a visible action "used as an utterance or as a part of an utterance" with the goal of communicating something. Gestures can be classified according to their communicative function, the kind of meaning they express, their co-occurrence with speech, and their degree of conventionality and arbitrariness, among other factors (McNeill 1992; Kendon 2004; McNeill 2005; Abner et al. 2015; among others). Other types of gestures we will not be addressing include *emblems* (culture-specific gestures with fixed, arbitrary meanings that can be used with or without accompanying speech, such as the "thumbs-up" gesture), *pantomimes* (sequences of complete gestures that iconically depict scenes and/or events and never co-occur with speech), *beats* (short and simple non-iconic movements that pattern closely with an utterance's prosodic peaks), and *deictics* (pointing gestures commonly used with demonstratives and locative adverbs). Iconic co-speech gestures can be distinguished from these other types of gestures by their co-occurrence with speech, their lack of arbitrariness, and their conveyance of meaning through iconicity. However, like other gestures, they are naturally occurring linguistic phenomena that are part of the linguistic system (Kendon 1980; McNeill 1985; 1992; Goldin-Meadow & Brentari 2017); they are spontaneously produced by speakers, and are not formally taught to learners.

The modest amount of previous formal semantic/pragmatic work on co-speech gestures has generally focused on inferences related to their not-at-issue-ness (Ebert & Ebert 2016; Tieu et al. 2017; Esipova 2018; Schlenker 2018), as illustrated by (7) where denial of the utterance cannot target the content of the gesture as in (b); instead (7a) can be continued by (b′) or (b″).

(7)   a.   John brought a [bottle of beer]_LARGE.
        b.   …#No, it was small.
        b′. …Yeah, but it was a small one.
        b″. …Yeah, and it was huge, you're right!

Of not-at-issue types of meaning, Schlenker (2018) proposes that co-speech gestures are in fact a type of presupposition, specifically an assertion-dependent conditional presupposition he calls a *cosupposition*. Based on informal inference judgments, he demonstrates that these cosuppositions exhibit standard presuppositional projection behavior in embedded contexts, except that the inference that projects is *conditional*. (In the examples below, reported inferences are indicated with the symbol ⇒.)

(8)   Projected cosuppositional inferences of co-speech gestures (Schlenker 2018: 3–13)
        a.   John [helped]_UP his son.
             ⇒ John helped his son by lifting him.
        b.   John didn't [help]_UP his son.
             ⇒ If John had helped his son, he would have done so by lifting him.
        c.   If little Johnny takes part in the competition, will his mother [help]_UP him?
             ⇒ If little Johnny takes part in the competition, if his mother helps him, lifting will be involved.
        d.   None of these ten guys [helped]_UP his son.
             ⇒ For each of these 10 guys, if he had helped his son, this would have involved some lifting.
        e.   Does Samantha believe that John [helped]_UP his son?
             ⇒? Samantha believes that if John helped his son, lifting was involved.
             ⇒?? If John helped his son, lifting was involved.

In contrast, Ebert & Ebert (2014; 2016) suggest that the way in which co-speech gestures contribute not-at-issue meaning is most analogous to *supplements*, such as expressives or non-restrictive relative clauses (Potts 2005). It is a subtle distinction: consider the case of non-restrictive relative clauses outlined in Chierchia & McConnell-Ginet (2000), which exhibit projection out of several operators in (9).

(9)   Chierchia & McConnell-Ginet (2000: 351)
   a.   Jill, who lost something on the flight from Ithaca to New York, likes to travel by train.
   b.   Jill, who lost something on the flight from Ithaca to New York, doesn't like to travel by train.
   c.   Does Jill, who lost something on the flight from Ithaca to New York, like to travel by train?
   d.   If Jill, who lost something on the flight from Ithaca to New York, likes to travel by train, she probably flies infrequently.

Sentence (9a) has the implication that Jill lost something on the flight from Ithaca to New York; sentences (b–d) show that the inference survives under negation, in a question, and in a conditional. However, unlike classic presuppositions, in none of (a–d) is it assumed by the speaker that the hearer already knows that Jill lost something, so these inferences have typically not been classified as presuppositions since they impose no restrictions on the background content.

An example of a supplemental analysis of the meaning of a co-speech gesture is given in (10), where Ebert & Ebert (2016) illustrate how the supplemental analysis predicts that the material conveyed by the gesture goes through as an inference in positive contexts (10a) but not in negative contexts (10b).

(10)   Supplemental inferences of co-speech gestures (Ebert & Ebert 2016)
   a.   Some philosopher brought [a bottle of beer]_BIG yesterday.
        ⇒ Some philosopher brought a bottle of beer, which was big.

   b.   #No philosopher brought [a bottle of beer]_BIG yesterday.
            (Intended inference: No philosopher brought a bottle of beer, which was big.)

Although both the cosuppositional analysis and the supplement analysis of co-speech gestures take them to be not-at-issue, they make different predictions in several respects, one of which is projection behavior under negation, shown in (10) above. We will focus on one previously unexplored contrast in this paper: the restrictions that the gestures place on the discourse, either in previous context or in the same sentence, specifically whether the gesture is *trivial*, duplicating content provided elsewhere. Supplements are awkward when trivial: compare (11a), where the supplement is trivial, to (11b), a similar structure but where the supplement contains nontrivial information.

(11)   a.   #My friend Jill lost her phone on her flight from Ithaca to New York yesterday.
             Jill, who lost something on the flight from Ithaca to New York, likes to travel by train.
        b.   My friend Jill lost her phone on her flight from Ithaca to New York yesterday.
             Jill, who frequently travels from Ithaca to New York, likes to travel by train.

Taking a supplement analysis of co-speech gesture off the shelf, we would expect degraded acceptability for co-speech gestures that are trivial, by analogy to speech supplements. The cosuppositional analysis makes a less overt prediction about triviality, but by analogy to presupposition we might in fact expect the reverse direction of acceptability: presuppositions

are usually expected to be given/trivial, and so the cosuppositional analysis would be consistent with the finding that gestures are more acceptable when trivial.

Past experimental work on the cosupposition/supplement distinction comes from Tieu et al. (2017), who report evidence based on different patterns of projection in favor of the cosuppositional analysis, albeit one with some extra assumptions that need to be made. We note, however, that their work focuses on *inferential* judgments involving gesture, while we take as our starting point the quite different task of *acceptability* judgments of gestures in context. Our motivation for focusing on acceptability is that we begin with the impression that, to the extent that we have acceptability judgments about co-speech gestures, they vary greatly in ways that we cannot predict using current theories of co-speech gesture. This includes examples in existing literature, some of which are entirely natural (e.g., (2)) and others which we found to be less so (e.g., (8d)), and which cannot in our mind yet be entirely explained by processing constraints like complexity, frequency, familiarity, etc. We are therefore interested in asking what kinds of factors contribute to the acceptability of co-speech gestures, so that gesture researchers creating linguistic examples can make more informed choices, including for important patterns like inferential judgments.

To summarize, we are interested in focusing on acceptability given the not-at-issue tendency of co-speech gestures, since dependence on previous context (in particular, being trivial) is unexpected for supplements but a hallmark of one type of presuppositional content (so-called "hard triggers"). With an eye toward fieldwork (or, in our case, work on an understudied phenomenon), Tonhauser et al. (2013) propose a typology for projective content. One of the properties that they note forms a class of projective content is the *strong contextual felicity (SCF)* condition, which will be the focus of this paper. By directly applying their suggested methodology for investigating expressions that seem to show projective behavior, we hope to better understand how to classify co-speech gestures in English in comparison to other projective spoken language phenomena, and to gain some more empirical evidence to bear on the theoretical analysis of co-speech gesture.

1.2 Background: Testing strong contextual felicity

The goal of our study is to essentially test how easily the content of not-at-issue meaning can be accommodated by an interlocutor. In other words, we are interested in asking whether co-speech gestures are more or less acceptable when they duplicate content or are entailed by content in the preceding discourse, or when they contribute new content. That is, are gestures better when they are informative, either with respect to the past context or the utterance that contains them? We will get at the former by testing whether their content must be entailed by the preceding discourse. Tonhauser et al. (2013) refer to this sensitivity to discourse context as *strong contextual felicity (SCF)*; more specifically, an item is said to be [+SCF] if and only if it is only acceptable in a discourse context that appropriately entails it, where context entailment is defined in (12):

(12)   Tonhauser et al. (2013: 75–76)
      a.   An *m-positive context* is one that entails or implies *m*.
      b.   An *m-neutral context* is one that entails or implies neither *m* nor ¬*m*.

Then the property of strong contextual felicity is defined as follows:

(13)   *Strong contextual felicity constraint* (Tonhauser et al. 2013: 76)
      If utterance of trigger *t* of projective content *m* is acceptable only in an *m*-positive context, then *t* imposes a *strong contextual felicity constraint with respect to m*.

Informally, if some expression can be uttered in a discourse context that neither supports nor denies an inference m, then that expression does not need to be supported by the context, so it is [–SCF] with respect to m. Formally, the following diagnostic can be used to decide on the strong contextual felicity value of an expression:

(14)   *Diagnostic for strong contextual felicity* (Tonhauser et al. 2013: 76)
        Let S be an atomic sentence that contains trigger t of projective content m.
        (i)   If uttering S is acceptable in an m-neutral context, then trigger t does not
              impose a strong contextual felicity constraint with respect to m.
        (ii) If uttering S is unacceptable in an m-neutral context and acceptable in a mini-
              mally different m-positive context, then trigger t imposes a strong contextual
              felicity constraint with respect to m.

Part (ii) of the diagnostic captures the intuition that if an expression is not felicitous in a neutral context, but then is felicitous in a nearly identical context in which the only difference is that now m is entailed by the context, then we can safely conclude that the expression is [+SCF]. Using this diagnostic, we devised an experiment using Amazon Mechanical Turk to test if the semantic content of co-speech gestures is [+SCF] or [–SCF] by constructing various scenarios that permit co-speech gestures and manipulating the context to be either m-neutral or m-positive, where m is the proposition conveying the semantic content of the co-speech gesture.

As an example, consider the English additive particle 'too', which is standardly analyzed as having an existence presupposition of a salient parallel alternative proposition.

(15)   Tonhauser et al. (2013: 78–79)
        a.   [Context: Malena is eating her lunch, a hamburger, on the bus going into
             town. A woman who she doesn't know sits down next to her and says:]
             #Our bus driver is eating empanadas, too.

        b.   [Context: Same as in (15a), but Malena is eating empanadas.]
             Our bus driver is eating empanadas, too.

Let m be the proposition that somebody besides the bus driver is eating empanadas. The context in (15a) is m-neutral, because it doesn't specify that anyone else is eating empanadas (Malena is eating a hamburger), or that no one else is eating empanadas. In this context, the sentence with *too* is infelicitous. The minimally different context in (15b), by contrast, is m-positive, since Malena is now said to be eating empanadas, not a hamburger. In this context, the sentence with *too* is felicitous. Since the same utterance with *too* is infelicitous in an m-neutral context but felicitous in a minimally different m-positive context, we conclude by the diagnostic in (14) that *too* introduces a strong contextual felicity constraint, i.e., *too* is [+SCF] with respect to the implication m.

An example of a projection trigger that does not exhibit a strong contextual felicity constraint is the change-of-state construction *stop X*.

(16)   Adapted from Tonhauser et al. (2013: 80)
        [Context: Laura, who doesn't live with her parents, visits them and asks them to
        sit down with her because she wants to tell them something:]
        I've stopped eating gluten.

Let m be the proposition that Laura used to eat gluten. The context in (16) is neutral with respect to m since Laura's parents are not asserted or implied to know about Laura's gluten consumption or lack thereof. Because the utterance is felicitous in the m-neutral

context, we conclude that *stop* does not impose a strong contextual felicity constraint, i.e., is [–SCF] with respect to the implication *m*.

Recall that we are interested in whether iconic co-speech gestures are better when they are *informative*, meaning either with respect to the past context or the utterance that contains them. We will address the second of these issues by adding an additional dimension to the SCF diagnostic, which is whether the semantic content of the gesture is duplicated in the same proposition. This is especially expected to bear on the analysis of co-speech gestures as supplements, which predicts that triviality/duplication should result in lower acceptability of co-speech gestures. Our hypotheses are as follows:

- Hypothesis regarding *strong contextual felicity*: A large category of expressions carrying not-at-issue content require that content to be entailed by the preceding context. We ask whether co-speech gestures have the same requirement.
- Hypothesis regarding matching *speech cue*: One hypothesized analysis of gestures is as supplements (Ebert & Ebert 2014), which are infelicitous when trivial. We ask how matching content in the speech stream affects felicity of co-speech gestures, with the expectation that it should decrease grammaticality under the supplement account.

We describe the methods and procedures of our experiment, including the implementation of both of these factors, in the next section.

2 Experiment 1: Co-speech gestures and context sensitivity

2.1 Methods

2.1.1 Participants

Participants were 198 adults recruited through Amazon Mechanical Turk, restricted to the United States region. All participants self-identified as native speakers of English. Participants were compensated monetarily for their participation in the questionnaire via Amazon payments of $2. Experimental protocols were approved by the Harvard University Institutional Review Board under approval number IRB16-1331.

2.1.2 Procedure

The questionnaire was created using the Qualtrics Survey Software platform. Amazon Mechanical Turk workers were directed to a Qualtrics link in order to complete the survey. The questionnaire took approximately 10–20 minutes to complete. Participants completed the survey on their own time and on a device of their choosing after receiving a link to the survey; they were instructed at the start to make sure to use a device with a large enough screen to play videos and with working speakers/headphones. The experimental task instructions given at the start of the survey were as follows:

> If you choose to be in the study, you will complete a questionnaire. This questionnaire will help us learn more about co-speech gestures in English. You will be asked to watch short video clips and judge the naturalness of English sentences, and **you may find all of them completely natural, all of them not completely natural, or a combination: some completely natural and some not completely natural**.

The term *co-speech gesture* was not defined and was mentioned nowhere else in the instructions. We also chose not to define the term *natural*, in order to not prejudice the participants toward judging only the spoken words or only the gestures. The experimental design (discussed in detail below) was between-subjects, ensuring that each participant

was shown either only videos with co-speech gestures or only videos without gestures, so as to help the gestures seem as natural as possible (since participants who saw gestures saw gestures in every trial). We worried that the participants viewing all no-gesture trials might feel uncomfortable rating every trial as "completely natural", which was the expected rating for non-gesture videos since all of the stimuli were designed to be felicitous English sentences. To counteract this possibility, we included and highlighted the bolded portion of the above instructions so that participants would know that this was a possible outcome and that they did not need to artificially rate some trials lower than others just for the sake of variety in their responses. (We will see in Experiment 2, Section 3, that the resulting pattern of judgments remained even after adding infelicitous non-gesture trials.)

For each of the 19 trials of the experiment, the participant was presented with the same instructions, namely to read the short context paragraph and then click "play" to view the following video. The context paragraphs were one- to two-sentence paragraphs describing a conversation between "Eliza", the speaker in the video, and another named interlocutor. Each video featured the same speaker, a female native English speaker in her early twenties who was identified as "Eliza" from the context paragraph, saying a sentence or two that continued the discourse started in the written context paragraph.

The embedded videos were hosted through YouTube and participants could replay the video if they wished, although no mention of the possibility of replaying videos appeared in the instructions. We did not collect data on the frequency of participants replaying videos.

Beneath the video, participants were prompted to make a binary naturalness judgment in response to the following prompt:

Please rate how natural you find Eliza's response in the video.

Participants were forced to press either the *completely natural* button or the *not completely natural* button to move on to the next trial. We specifically chose to describe Eliza's utterance (speech and gesture) in the video as a *response* because this speech word seemed to best encompass both her spoken words and her gestures. Other alternatives such as *utterance, statement, performance,* etc. we deemed to be either too technical or to implicitly bias the participants towards judging only the spoken words or only the gestural production. However, the term *response* implies that Eliza is part of a dialogue; hence we intentionally wrote each pre-video context paragraph as a dialogue, either implicit or explicit, between Eliza and another named interlocutor.

Each participant was presented with three "attention check" trials randomly interspersed between the experimental trials. These attention checks consisted of a context paragraph, written in the same manner as the experimental contexts, and a video in which Eliza informs the participant that this is an attention check and that they should press a particular response and move on to the next trial. One of the attention checks instructed the participant to choose "completely natural"; the other two instructed them to choose "not completely natural". These attention checks were used to filter out responses from participants who were not watching or not paying sufficient attention to the videos.

2.1.3 Stimuli

For each trial of the survey, participants saw a screen like that shown in Figure 1. At the top of the screen were instructions to read the context paragraph and then view the embedded video; below the video was the linguistic naturalness judgment question, along with the two options *completely natural* and *not completely natural*. Participants were forced to choose one or the other in order to proceed to the next trial.
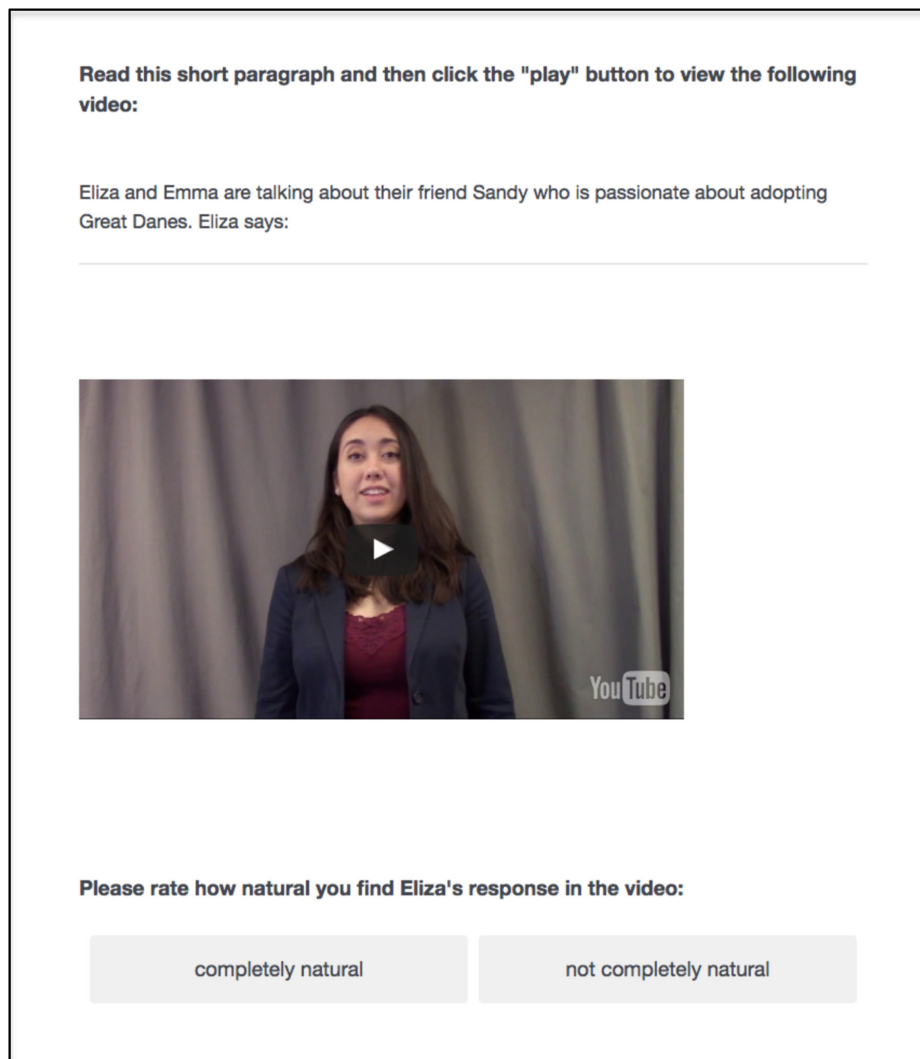
Figure 1: A screenshot of a trial, as seen by a participant.

Example (17) is one of the experimental *scenarios* seen by participants. The topic of the scenario is that Eliza and a friend are discussing a co-worker's wealth and jewelry. The participant would see either (i) or (ii) as the written context paragraph at the top of the page, and then watch a video with Eliza uttering one of (a)–(d). Utterances (a) and (b) contain co-speech gestures, while (c) and (d) do not. The gesture EARRING conveys the information that the type of diamond jewelry Alicia was wearing was earrings; this information is expressed under *Proposition m* in the example below.

(17) Jewelry scenario, Experiment 1 (Scenario 9, Appendix A)
  (i) *m-neutral context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has a lot of money. Eliza agrees and says:
  (ii) *m-positive context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has a lot of money based on her new pair of earrings. Eliza agrees and says:
  (iii) *Proposition m for the gesture* EARRING: The type of jewelry was earrings.
    a. Alicia was wearing real diamond [jewelry]_EARRING at work this morning.
    b. Alicia was wearing real diamond [earrings]_EARRING at work this morning.
    c. Alicia was wearing real diamond jewelry at work this morning.
    d. Alicia was wearing real diamond earrings at work this morning.

In the appendices, we list the inference licensed by the co-speech gesture in each scenario explicitly as a proposition, so that it is clear what proposition *m* the *m-positive context* is supposed to entail/imply, in the Tonhauser et al. (2013) terminology.

In choosing co-speech gestures for the experiment, we avoided gestures that have conventionalized or codified meanings, such as a "thumbs-up" gesture, as these fall under the category of "emblems" and do not need to be accompanied by speech/sign to convey meaning (Kendon 2004). We also chose not to include pointing gestures, as these presumably intersect non-trivially with the semantics of deictic expressions (e.g., see the treatment in Lascarides & Stone 2009). The choice of gesture target (i.e., the NP or VP semantically modified by the gesture) turned out to be as important as the choice of gesture in designing the stimuli; in order for a co-speech gesture to contribute non-trivially to the truth conditions of an utterance, the gesture target needs to be semantically underspecified in some way. Modeling our examples after those discussed in Ebert & Ebert (2016) and Schlenker (2018), we chose gesture targets that are underspecified in manner when the target is a VP, or in adjectival content when the target is an NP; hence the co-speech gestures in our experiment function semantically as either manner adverbials or adjectival modifiers.

Figure 2 shows screencaps (captured "mid-action") of six of the co-speech gestures appearing in the videos shown to participants who were assigned to the "gesture-is-present" condition. These six co-speech gestures correspond to Eliza's video utterances like the following:

(18)   a.   Scenario 2 (Appendix A)
              The basketball match she was in last night was incredible! She [scored]_
              SHOOT the winning point!

        b.   Scenario 3 (Appendix A)
              Sandy just got [a dog]_BIG yesterday, and I hear it's quite the handful!

        c.   Scenario 6 (Appendix A)
              The moon was so gorgeous last night — we just sat outside [looking up]_
              TELESCOPE at it for awhile.

        d.   Scenario 7 (Appendix A)
              Alex kept [checking the time]_WRISTWATCH during the date.

        e.   Scenario 9 (Appendix A)
              Alicia was wearing real diamond [earrings]_EARRING at work this morning.

        f.   Scenario 10 (Appendix A)
              Lisa [performed]_VIOLIN really well at the recital last night!

Figure 2: Sample screencaps of co-speech gestures.

For a full listing of the written and spoken experimental stimuli for Experiment 1, see Appendix A. Appendix B provides screencaps for all 16 co-speech gestures appearing in the videos.

2.1.4 Design

Each participant saw 19 trials total: 16 experimental trials and 3 attention check trials. Each experimental trial (*scenario*) came in one of 8 types, depending on (i) whether or not there was a co-speech gesture in the video (the GESTURECUE condition); (ii) whether or not there was a *speech cue* in the video, i.e., a linguistic expression verbally expressing the content of the gesture (the SPEECHCUE condition); and (iii) whether the video was presented with a *neutral* context paragraph or a *positive* context paragraph (the CONTEXT condition), according to the Tonhauser et al. (2013) definitions; see Table 1 for a list of all eight trial types. We discuss the three experimental factors in more detail below.

At the start of each survey, the participant was randomly assigned to either the [GESTURECUE = yes] group or the [GESTURECUE = no] group. The sixteen scenarios were then presented in a randomized order (with the three attention checks randomly interspersed). For each scenario, a 2 × 2 Latin square design was used which crossed the two factors CONTEXT and SPEECHCUE. Each participant saw a trial for every scenario.

The GESTURECUE condition. The GESTURECUE condition encodes the presence ("yes") or absence ("no") of a co-speech gesture in the video utterance. As described above, this condition was manipulated between subjects; each participant saw either only videos with co-speech gestures or only videos without co-speech gestures. This design was chosen so as to help the gestures seem as natural as possible and to elicit more subtle felicity judgments, since there seemed to us to be a very real possibility that participants would rate any video with a gesture as worse than a video without one (as indeed turned out to be the case).

Within a scenario and given identical SPEECHCUE values, the corresponding videos with and without co-speech gestures were identical; that is, the words spoken by Eliza were exactly the same and were delivered, as nearly as possible, with the same intonation.

In trials with co-speech gestures, one factor that was not systematically controlled for was whether the gesture modified (i.e., whether the gesture *target* was) a verb or a noun. Consider the following two examples:

(19) a. Scenario 10 (Appendix A)
          Lisa [performed]_VIOLIN really well at the recital last night!

      b. Scenario 3 (Appendix A)
          Sandy just got [a dog]_BIG yesterday, and I hear it's quite the handful!

Table 1: Experiment 1 trial types.

Trial type	GESTURECUE	SPEECHCUE	CONTEXT
1	no	no	neutral
2	no	no	positive
3	no	yes	neutral
4	no	yes	positive
5	yes	no	neutral
6	yes	no	positive
7	yes	yes	neutral
8	yes	yes	positive

In (19a), the gesture target *performed* is a verb, while in (19b) the gesture target is the NP *a dog*. Out of the 16 total experimental scenarios, 13 had gestures modifying VPs, while only 3 had gestures modifying NPs, and one of these three needed to be excluded from analysis due to an accidental name mismatch between the contexts and the videos (Scenario 11). We discuss the results of a follow-up analysis on the data based on this division (noun/verb) in Section 2.3.

The SPEECHCUE condition. The SPEECHCUE condition encodes the presence ("yes") or absence ("no") of a spoken expression in the video that (approximately[1]) duplicates the same semantic content as the corresponding co-speech gesture for that scenario. We refer to this spoken linguistic expression as a *speech cue*. Consider the following pairs of target utterances (20) and (21), in which the (a) utterances do not have a speech cue (SPEECHCUE = no), while the (b) utterances do (SPEECHCUE = yes); the speech cues in question are bolded in the (b) utterances:

(20)   Scenario 2 (Appendix A)
        a.   The match she was in last night was incredible! She [scored]_SHOOT the
              winning point!
        b.   The **basketball** match she was in last night was incredible! She [scored]_
              SHOOT the winning point!

(21)   Scenario 3 (Appendix A)
        a.   Sandy just got [a dog]_BIG yesterday, and I hear it's quite the handful!
        b.   Sandy just got [a **big** dog]_BIG yesterday, and I hear it's quite the
              handful!

In (20b) the speech cue *basketball* indicates that the sporting event described is a basketball game, and the co-speech gesture SHOOT depicts the act of shooting a basketball; hence both contribute the semantic information that the sport in question is basketball (and so the co-speech gesture is *trivial* in (20b)). Similarly, in (21b) the speech cue *big* indicates that Sandy's dog is a big dog, and the co-speech gesture BIG manually depicts the same information (and is hence trivial).

   Although the speech cues duplicate the semantic content of the designated co-speech gesture for a given scenario, in our stimuli a speech cue can and does appear in trial types for which [GESTURECUE = no], i.e., in video utterances that do not have a gesture. In these trial types, the speech cue is exactly the same as that which appears in the utterance that has the co-speech gesture for that scenario.

   The timing of the SPEECHCUE with respect to its (roughly) equivalent co-speech gesture varied across scenarios. In some scenarios, the speech cue phrase occurred before the gesture target (22); in others, it was contained within or was the entire gesture target phrase (23); and in still others it occurred after the gesture target (24). (As before, in the following examples the speech cues are bolded for identification purposes.)

(22)   Scenario 6 (Appendix A)
        a.   The moon was so gorgeous last night — we just sat outside [looking up]_TEL-
              ESCOPE at it for awhile.
        b.   The moon was so gorgeous last night — we just sat outside **and took turns
              with the telescope** [looking up]_TELESCOPE at it for awhile.

---

[1] The content of the SPEECHCUE and the co-speech gesture are not strictly equivalent, of course, given the iconic, analog nature of the co-speech gesture.

(23)   Scenario 4 (Appendix A)
    a.   Karen [ran]_RUN-DOWN to see what was happening!
    b.   Karen [ran **down**]_RUN-DOWN to see what was happening!

(24)   Scenario 7 (Appendix A)
    a.   Alex kept [checking the time]_WRISTWATCH during the date.
    b.   Alex kept [checking the time]_WRISTWATCH **on his watch** during the date.

In (22b), the bolded speech cue *and took turns with the telescope* conveys that the looking up was done with a telescope, just as the gesture TELESCOPE does, and it occurs linearly before the gesture target *looking up*. In (23b), on the other hand, the bolded speech cue *down* occurs within the gesture target verb phrase *run down*. Finally, in (24b), the bolded speech cue *on his watch* occurs immediately after the gesture target and, like the gesture WRISTWATCH, conveys that Alex checked the time using his wristwatch.

As discussed above, the SPEECHCUE condition was manipulated within subjects and crossed with CONTEXT in a 2 × 2 Latin square design.

The CONTEXT condition. The CONTEXT condition encodes whether the written paragraph shown to participants before the video entails or implies the content of the co-speech gesture (a *positive* context), or whether it neither entails/implies the content of the co-speech gesture, nor entails/implies its negation (a *neutral* context). The "positive" and "neutral" terminology follows the naming conventions for *m*-neutral and *m*-positive contexts in Tonhauser et al. (2013), where here *m* is the semantic proposition expressing the information conveyed by the co-speech gesture. The effect is to manipulate the informativity/triviality of gestural content. An example can be seen in (25); the minimal differences between the two contexts are bolded for identification purposes.

(25)   Scenario 2 (Appendix A)
    a.   *Neutral context*: Eliza and Tom are talking about the Olympics, and Eliza is telling Tom about her favorite new **athlete** who Tom hasn't heard of. Eliza says:
    b.   *Positive context*: Eliza and Tom are talking about the Olympics, and Eliza is telling Tom about her favorite new **basketball player** who Tom hasn't heard of. Eliza says:
    c.   *Sample target utterance*: The match she was in last night was incredible! She [scored]_SHOOT the winning point!

For this scenario, participants assigned [CONTEXT = neutral] were shown the context paragraph in (25a), while participants assigned [CONTEXT = positive] were shown the context paragraph in (25b). As seen in the sample target utterance (25c), the co-speech gesture for this scenario is SHOOT, which conveys that the scoring event was done by shooting a basketball. The "neutral" context (a) does not specify what type of sporting event is being discussed; hence it entails/implies neither SHOOT nor ¬SHOOT, meeting the criteria of being an *m*-neutral context. By contrast, the "positive" context (b) contains the information that the athlete in question is a basketball player; this entails that were this athlete to score in a game, she would do so by shooting a basketball. This is exactly the content of the gesture SHOOT, so (b) entails the content of the co-speech gesture and meets the criteria of being an *m*-positive context.

Notice that the positive and neutral context paragraphs in (25) are extremely similar; they differ only in the replacement of the phrase *athlete* with *basketball player*. In general we kept the neutral/positive context pairs as minimally different as possible, only making those changes from the neutral to the positive that were necessary to entail or imply

the proposition expressed by the co-speech gesture. This follows Tonhauser et al.'s (2013) recipe to have an *m*-positive context be minimally different from an *m*-neutral context, in order to tell if *m* does indeed require the support of the context to be used felicitously.

The CONTEXT condition was manipulated within subjects, and, as discussed above, was crossed with SPEECHCUE in a 2 × 2 Latin square design.

2.2 Results

Out of 198 participants, 5 participants' responses were excluded due to those participants failing at least one of the three attention checks. The results discussed below are for responses from the remaining 193 participants.

Recall that the experimental design included eight trial types resulting from all possible combinations of the experimental conditions GESTURECUE, SPEECHCUE, and CONTEXT. Table 2 reports the mean acceptance rates, standard deviations, and standard errors across the eight trial types, and Figure 3 shows mean acceptance rates for each trial type with standard error bars. Descriptively, we see an interesting pattern in which the presence of a co-speech gesture decreases acceptability (as indicated by the values of M for trial

Table 2: Experiment 1 trial types with conditions (GESTURECUE, SPEECHCUE, and CONTEXT) and mean acceptance rates.

Trial type	GESTCUE	SPCUE	CONTEXT	M	SD	N	SE
1	no	no	neut	0.806	0.396	371	0.021
2	no	no	pos	0.775	0.418	360	0.022
3	no	yes	neut	0.771	0.421	363	0.022
4	no	yes	pos	0.734	0.442	361	0.023
5	yes	no	neut	0.607	0.489	364	0.026
6	yes	no	pos	0.604	0.490	359	0.026
7	yes	yes	neut	0.712	0.454	361	0.024
8	yes	yes	pos	0.691	0.463	356	0.025
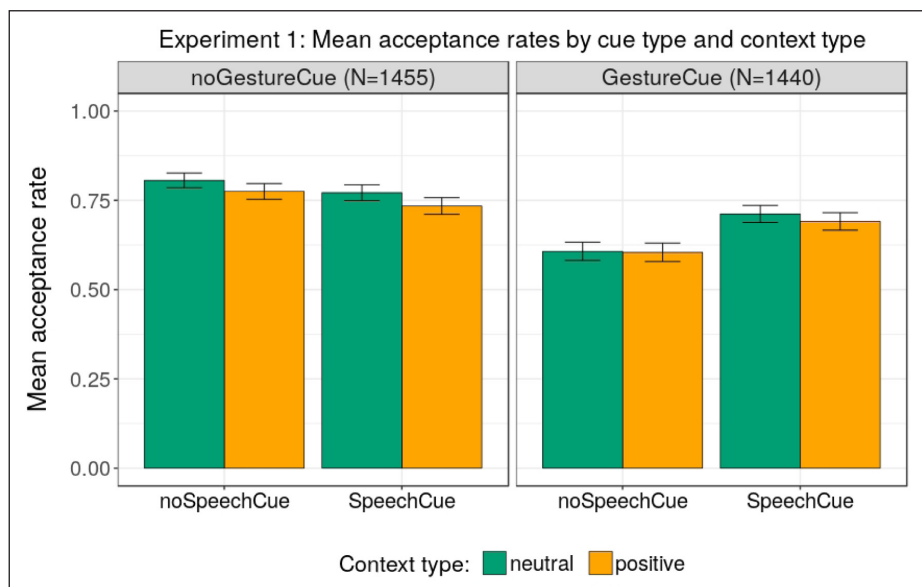
Figure 3: Experiment 1 mean acceptance rates, by experimental conditions GESTURECUE, SPEECHCUE, and CONTEXT.

types 5–8), although the presence of a matching speech cue with the gesture somewhat mitigates this effect (as indicated by the values of M for trial types 7 and 8).

With respect to the experimental conditions GESTURECUE, SPEECHCUE, and CONTEXT, trends were as follows. On average, trials with gestures (trial types 5–8) were rejected more often than trials without gestures (trial types 1–4) (with gesture: M = 0.65; without gesture: M = 0.77). Trials shown with a neutral context (1, 3, 5, and 7) were rejected at approximately the same rate as trials shown with a positive context (2, 4, 6, and 8) (neutral: M = 0.72; positive: M = 0.70). Finally, trials with a speech cue (3, 4, 7, and 8) were rejected on average at approximately the same rate as trials without a speech cue (1, 2, 5, and 6) (with speech cue: M = 0.73; without speech cue: M = 0.70).

Table 3 gives a breakdown of ratings by scenario. Scenario 11 trials were excluded from analysis because of an accidental name mismatch between the written contexts and the speech in the video. Means for the remaining 15 scenarios ranged from 0.57 (Scenario 1) to 0.82 (Scenario 5), with SDs ranging from 0.38 to 0.50.

Analyses of subjects' judgment responses were conducted using the R programming language (version 3.2.3) (R Core Team 2016) to build generalized linear mixed effects models (Baayen et al. 2008), using the function *glmer* from the *lme4* package (Bates et al. 2015). In the model with the most data variation coverage (determined by ANOVA testing), the three independent factors were GESTURECUE (no/yes), SPEECHCUE (no/yes), and CONTEXT (neutral/positive); the binary naturalness judgment (coded as 1 = "ompletely natural" and 0 = "not completely natural") was the dependent variable, and scenario number and participant ID were coded as random effects.

Results of the model indicate that there was a significant main effect of GESTURECUE (β = –1.148, z = –4.908, p < 0.001), but also a significant interaction of the factors GESTURECUE and SPEECHCUE (β = 0.665, z = 2.507, p < 0.05). Neither SPEECHCUE nor CONTEXT were significant main effects (p > 0.1); these findings agree with other models of the data that we constructed in which the only fixed factor was SPEECHCUE or CONTEXT (p > 0.1 in each case). There were no other significant interactions between factors.

2.3 Discussion

The main experimental question of this study was how contextual support (within a discourse or the same utterance) influences the felicity of co-speech gestures. The first factor we tested was whether co-speech gestures need to be entailed by their preceding discourse context to be used felicitously. In the terminology of Tonhauser et al. (2013), we wanted to find out if co-speech gestures are [+strong contextual felicity (SCF)] or [–SCF]. Our experimental results indicate that manipulating the discourse context from being *m*-neutral to *m*-positive (where *m* is the proposition representing the semantic content of the co-speech gesture) had no significant effect on participants' acceptance rates of stimuli items. Since participants were equally likely to judge an utterance as completely natural with or without the positive entailment of the content of the co-speech gesture, and assuming that participants were judging the video with the context paragraph in mind, we can tentatively conclude that co-speech gestures are [–SCF]. In other words, co-speech gestures can either be easily accommodated or need no contextual entailment to be used felicitously. This is in clear contrast to other projective content like the English additive

Table 3: Experiment 1 acceptance rates by scenario.

Scen	1	2	3	4	5	6	7	8	9	10	12	13	14	15	16
M	.57	.66	.77	.70	.82	.75	.74	.74	.81	.75	.78	.66	.65	.69	.62
SD	.50	.47	.42	.46	.38	.44	.44	.44	.40	.44	.42	.48	.48	.46	.49

particle *too*, a so-called "hard" presupposition trigger (Abrusán 2016); as we discussed in example (15) in Section 1.2, *too* is [+SCF] with respect to its implication of the existence of a salient parallel alternative proposition in the discourse context. The differing behavior of *too* and co-speech gestures with respect to the content of the contexts they appear in supports the claims that the former is [+SCF] (Tonhauser et al. 2013) and the latter is [–SCF]. In sum, our experiment has shown that co-speech gestures cannot be classified as "hard" presupposition triggers (assuming participants did indeed read and include the context paragraphs in their linguistic judgments, which concern we will address with Experiment 2 in Section 3).

The significant main effect of GESTURECUE in the data indicates that the presence of a co-speech gesture in a trial in general had a negative effect on the rating of the video. Intriguingly, this negative effect was mitigated when there was a speech cue present along with the gesture in the trial. One way to interpret this is that gestures are used most felicitously when they are semantically duplicating information already present in the speech/sign stream. This redundancy preference clearly pragmatically differentiates co-speech gestures from well-studied non-gestural supplemental material like appositives; appositives are not felicitous when they convey the same information as the main assertion (Potts 2005), as can be seen in (26):

(26)   Adapted from Ebert & Ebert (2016)
        Paul, [the best horse riding instructor in the world], moved to Stuttgart recently
        (#and is the best horse riding instructor in the world).

The content of the appositive NP in (26), shown in brackets, cannot be reiterated as part of the foreground asserted content. However, as we see from the experimental results, partial and even complete redundancy is completely acceptable with co-speech gestures; (27) is an example of a trial from the experiment with both a co-speech gesture and a speech cue present (Scenario 3, Appendix A).

(27)   Sandy just got [a **big** dog]_BIG yesterday, and I hear it's quite the handful!

Here the gesture BIG is arguably contributing very similar content to the speech cue *big*.

The significant interaction between GESTURECUE and SPEECHCUE indicates that speech cues had a different effect on judgments depending on whether or not there was a gesture; namely, trials with gestures had a higher average acceptance rate with a speech cue than without, while trials without gestures had a lower average acceptance rate with speech cues than without. Presumably this has something to do with pragmatic calculations about the amount of information present in the utterance and any corresponding implications the hearer (our participants) might have drawn. Clearly, at this point our results raise more questions than answers, especially regarding whether the same pattern would emerge with different kinds of co-speech gestures.

Since there was such wide variation in ratings across scenarios, we did several follow-up analyses (which were not planned before the experiment was run) of scenario types to see how properties that varied by scenario affected judgments. We focused on three questions: (i) whether the gesture semantically modified a noun or verb; (ii) whether the speech cue that duplicated the gesture's content came temporally before, during, or after the co-speech gesture; and (iii) whether the belief state of the addressee, given the context paragraph, affected judgments. For the first question, we found that, overall, co-speech gestures that appeared in scenarios where they were modifying nouns were more acceptable than those modifying verbs (**nouns:** *neutral context*: M = 0.83, SD = 0.38; *positive context*: M = 0.77, SD = 0.42; **verbs:** *neut*: M = 0.634, SD = 0.48; *pos*: M = 0.628, SD = 0.48). However,

the number of scenarios is too small to generalize beyond the sample (gestures modified nouns in two scenarios and verbs in thirteen scenarios); see Figure 4.

Next, speech cues that occurred after their corresponding gesture were generally less acceptable than those that occurred with or before the gesture (**after the gesture** (N = 238): *neutral context*: M = 0.63, SD = 0.49; *positive context*: M = 0.65, SD = 0.48; **before the gesture** (N = 93): *neut*: M = 0.78, SD = 0.42; *pos*: M = 0.75, SD = 0.44; **during the gesture** (N = 386): *neut*: M = 0.75, SD = 0.43; *pos*: M = 0.70, SD = 0.46), but again, the sample size of each type of scenario is small (speech cues occurred after the gesture in five scenarios, before the gesture in two scenarios, and with the gesture in eight scenarios); see Figure 5.
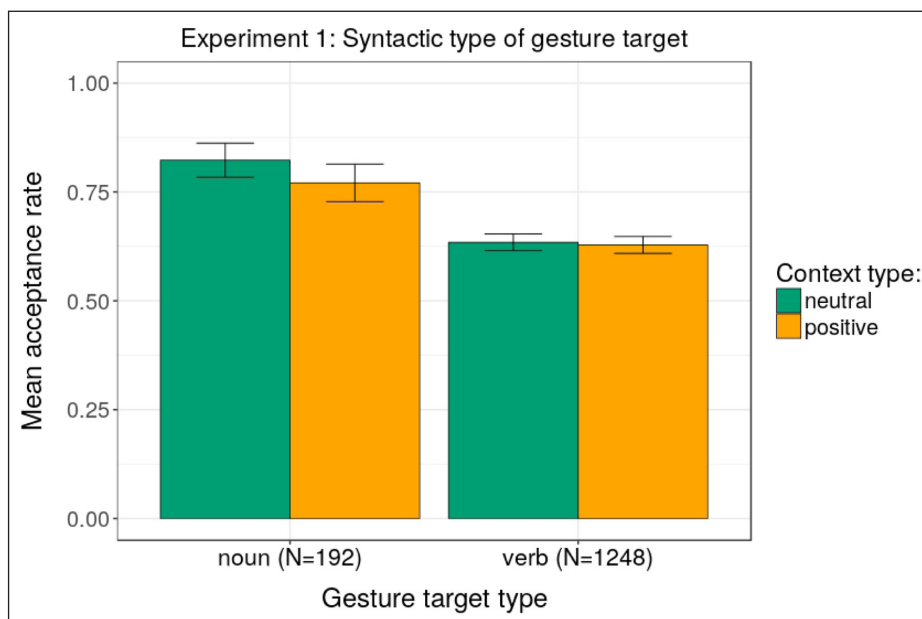
Figure 4: Experiment 1 mean acceptance rates for type of syntactic item modified by gesture (noun vs. verb), by context type.
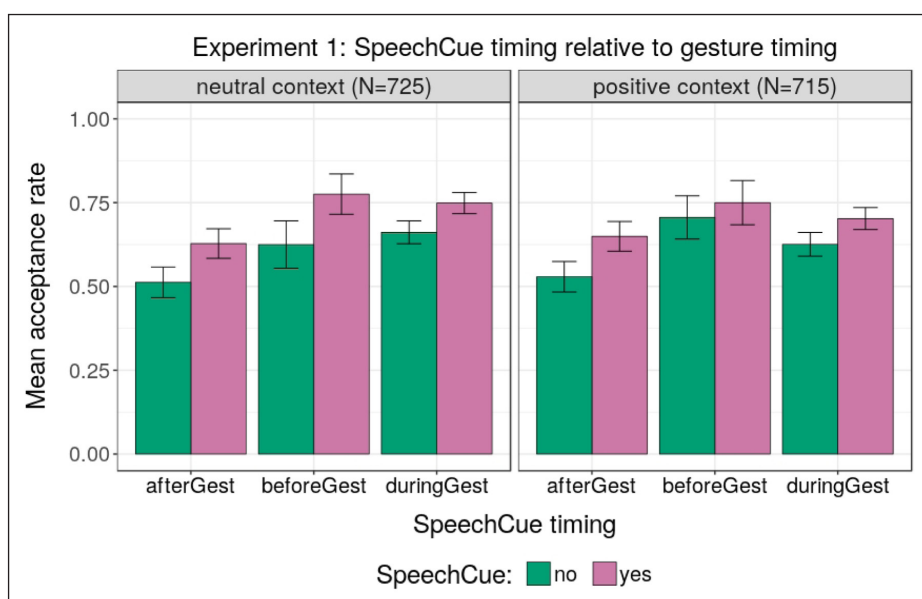
Figure 5: Experiment 1 mean acceptance rates for speech cue timing relative to gesture timing, by context type.

The differences between placement of speech cues (before, during, and after the gesture) may initially appear to be intriguing, but critically there was no difference between these timing conditions for speech cues in trials where the speech cue was present versus those where it was absent. Using the *glmer* function in R, we constructed a generalized linear mixed effects model (fit by Maximum Likelihood) on the subset of the data where a gesture was present [Gesture Cue = yes], with subject ID and scenario number as random effects, the participant's naturalness judgment (Response) as the dependent variable, and SpeechCue (yes/no) and SpeechCueTiming (before/during/after) as the independent variables. Neither the interaction between the presence of a speech cue and the speech cue coming after the gesture, nor the interaction between the presence of a speech cue and the speech cue coming before the gesture, was significant ($p > 0.1$ in both cases). This suggests that the difference in ratings between the differing speech cue positions is an artifact of the large variation in trial types and not in the end about the speech cues themselves.

Another factor one might suspect of influencing acceptability judgments by participants is the belief state of the addressee in the "dialogue" between Eliza and her interlocutor described in the context paragraphs. Does the addressee need to know/believe/be able to infer the content of the gesture proposition in order for the experimental participant to judge Eliza's use of the gesture felicitous?[2] We conducted a follow-up analysis on the data, classifying scenarios (yes/no) according to whether the addressee of the dialogue knows or could infer the content of the proposition in question based on the information provided by the "positive" context. We constructed two separate generalized linear mixed effects models, one on the "yes" subset of the data (where the addressee knows/can infer the semantic content of the gesture based on the positive context), and the other on the remaining "no" subset of the data, with Response as the dependent variable, SpeechCue, GestureCue, and Context as the independent variables, and subject ID and scenario number coded as random effects. Neither model showed a significant effect of Context on Response ($p > 0.1$). From this we conclude that the belief state of the addressee is not a significant factor in acceptability judgments of utterances with co-speech gestures based on prior discourse contexts.

In this experiment, there were several methodological issues which may have had an unintended effect and require follow-up studies. For one, we told participants that they were going to be asked to rate English sentences as *completely natural* or *not completely natural*; however, we did not explain or define the concept of linguistic naturalness. This may in fact have resulted in participants performing different rating tasks from each other due to interpreting "naturalness" in different ways. Some respondents, for example, may have been rating the "acting" performance of the speaker in the videos, rather than rating the felicity of the utterance (speech + gesture) alone. One way to address this might be to provide a simple, non-technical explanation of linguistic naturalness in the survey instructions so as to better communicate to participants what we as researchers are asking them to do. However, at this stage we are still unsure what exactly this would look like, and how one could dissociate performance factors from gesture with a prompt.

A more worrying methodological issue is that from Experiment 1, it is impossible to tell if participants were actually taking the time to read the written context paragraphs and if the contexts factored into their judgments of the video utterances. In fact this is related to our previous concern, because participants might rate the delivery (rather than the linguistic naturalness of the utterance) precisely because they weren't taking context into account. The attention checks built into the experiment could only detect if a participant

---

[2] We thank an anonymous reviewer for raising this interesting question.

was not watching or paying proper attention to the videos. Without knowing more about the participants' behavior with respect to the written contexts, we cannot fully conclude whether or not co-speech gestures need to be entailed by the discourse context. We address this concern in a follow-up experiment, Experiment 2, in the next section.

3 Experiment 2: Context follow-up

In order to address the methodological concern that participants were not reading the context paragraphs and/or not taking them into consideration in their rating of the videos, we conducted a follow-up study on MTurk in which participants were shown contexts that were pragmatically infelicitous with the video utterance, in addition to the usual neutral and positive contexts. The prediction was that if participants were truly reading the contexts and considering them when rating the videos, then the trials with infelicitous contexts would be accepted at a significantly lower rate. This prediction was indeed borne out.

3.1 Methods

3.1.1 Participants

Participants were 90 adults recruited through Amazon Mechanical Turk, restricted to the United States region. All participants self-identified as native speakers of English. Participants were compensated monetarily for their participation in the questionnaire via Amazon payments of $2. Experimental protocols were approved by the Harvard University Institutional Review Board under approval number IRB16-1331.

3.1.2 Procedure

The questionnaire was created using the Qualtrics Survey Software platform. Amazon MTurk workers were directed to a Qualtrics link in order to complete the survey. The questionnaire took approximately 10–20 minutes to complete. Participants completed the survey on their own time and on a device of their choosing after receiving a link to the survey; they were instructed at the start to make sure to use a device with a large enough screen to play videos and with working speakers/headphones. The experimental task instructions given at the start of the survey were the same as for Experiment 1 (see Section 2.1.2).

For each trial of the experiment, the participant was presented with the same instructions, namely to read the short context paragraph and then click "play" to view the following video. The video stimuli used were a subset of the video stimuli from Experiment 1. The embedded videos were again hosted through YouTube.

One significant difference between Experiment 1 and this follow-up experiment was the judgment task instructions shown to participants. In this study, beneath the video, participants were shown the following prompt:

> Please rate how natural you find Eliza's response in the video, given the context paragraph:

Participants again had to press either the *completely natural* button or the *not completely natural* button to move on to the next trial. The addition of the words …*given the context paragraph* to the prompt was intended to signal to participants that they should rate the video utterance in the context of the written paragraph (our concern from Experiment 1).

As in Experiment 1, each participant was presented with three "attention check" trials randomly interspersed between the experimental trials. As before, these attention checks were used to filter out responses from participants who were not paying sufficient attention to the videos.

3.1.3 Stimuli

The stimuli used for this follow-up study were all re-used from Experiment 1, except for the new context paragraphs of the "infelicitous" flavor (see below).

As an example, we return to the "jewelry scenario" in (28) (compare to example (17) from Experiment 1). In this new experiment, the participants saw one of *three* possible context variants (i, ii, or iii) as a written paragraph at the top of the page. Below this they saw the embedded video, which either contained a co-speech gesture (28a) or did not (28b). Finally, below the video were the judgment task instructions and the *completely natural* and *not completely natural* buttons. Infelicity was designed not to be egregious but frequently involved, e.g., a switch in names, as in (28iii) below. All of the stimuli used in Experiment 2 can be found in Appendix C.

(28)   Jewelry scenario, Experiment 2 (Scenario 9, Appendix C)
    (i)   *Neutral context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has a lot of money. Eliza agrees and says:
    (ii) *Positive context*: Eliza and Nina are gossiping about their coworker Alicia, and Nina says that she thinks Alicia has a lot of money based on her new pair of earrings. Eliza agrees and says:
    (iii) *Infelicitous context*: Eliza and Alicia are gossiping about their coworker Nina, and Alicia says that she thinks Nina has a lot of money. Eliza agrees and says:
        a.   Alicia was wearing real diamond [jewelry]_EARRING at work this morning.
        b.   Alicia was wearing real diamond earrings at work this morning.

Note that in this study, the experimental condition SPEECHCUE manipulated in Experiment 1 was eliminated since we only wanted to focus on the relationship between the context paragraphs and the rating of the video utterance. For the [GESTURECUE = yes] condition, we chose to re-use the "gesture-only" videos from Experiment 1, i.e., those that didn't have a supporting speech cue, as in (28a). By contrast, for the [GESTURECUE = no] condition, we picked the "speech cue-only" videos that did contain a speech cue, as in (28b).[3] This ensures that, for a given scenario, each member of the pair of videos (e.g., (28a,b)) conveys (roughly) the same overall semantic content, without duplication (again with the caveat that the two are not strictly equivalent, given that the gesture is iconic and interpreted in an analog way).

3.1.4 Design

Exactly as in Experiment 1, each participant saw 19 trials total: 16 experimental trials and 3 attention check trials. There were six distinct trial types, depending on (i) whether or not there was a co-speech gesture in the video (the GESTURECUE condition); and (ii) whether the video was presented with a "neutral" context paragraph, a "positive" context paragraph, or an "infelicitous" context paragraph (the CONTEXT condition) (Table 4). We discuss the two experimental conditions in more detail below.

At the start of each survey, the participant was randomly assigned to either the [GESTURECUE = yes] group or the [GESTURECUE = no] group. The sixteen scenarios were then presented in a randomized order (with the three attention checks randomly interspersed), and for each scenario the participant was randomly assigned one of the three CONTEXT condition values. Each participant saw a trial for every scenario.

---

[3] The former condition corresponds to Experiment 1 trial types 5 and 6, and the latter to trial types 3 and 4; see Table 1.

Table 4: Experiment 2 trial types.

Trial type	GESTURECUE	CONTEXT
1	no	infelicitous
2	no	neutral
3	no	positive
4	yes	infelicitous
5	yes	neutral
6	yes	positive

The GESTURECUE condition. As in Experiment 1, the GESTURECUE condition encodes the presence ("yes") or absence ("no") of a co-speech gesture in the video utterance. This condition was manipulated between subjects, for the same reasons discussed in Section 2.1.4 above for Experiment 1; thus each participant saw either only videos with co-speech gestures or only videos without co-speech gestures.

The CONTEXT condition. The CONTEXT condition was manipulated within subjects and encodes whether the written paragraph shown to participants before the video either (i) neither entails/implies the content of the co-speech gesture, nor entails/implies its negation (a *neutral* context), (ii) entails/implies the content of the gesture (a *positive* context), or (iii) is written so that the video utterance is a pragmatically infelicitous continuation of the context paragraph (an *infelicitous* context). Example (29) below shows the three context variants seen by participants for Scenario 5; minimal differences between contexts have been boldfaced for ease of comparison.

(29)   Scenario 5 (Appendix C)
   a.   *Neutral context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza spots her **there**, comes back to Jamie, and says:
   b.   *Positive context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza spots her **there on her computer**, comes back to Jamie, and says:
   c.   *Infelicitous context*: Eliza and Jamie are looking for Julia to join them for coffee. Jamie asks Eliza to check for her in the library. Eliza **doesn't see her there**, comes back to Jamie, and says:
   d.   *Sample target utterance*: I saw Julia over in the library [writing an essay]_ TYPE — it looks like she's a little preoccupied right now.

For this scenario, participants assigned [CONTEXT = neutral] were shown the context paragraph in (29a), participants assigned [CONTEXT = positive] were shown (29b), and participants assigned [CONTEXT = infelicitous] were shown (29c). As seen in the sample target utterance (29d), the co-speech gesture for this scenario is TYPE, which conveys that the essay-writing event was done by typing on a computer keyboard (as opposed to, e.g., with a paper and pencil). The "neutral" context (a) does not specify anything about Jamie's actions other than her location in the library; hence it entails/implies neither TYPE nor ¬TYPE, meeting the criteria of being an *m*-neutral context. By contrast, the "positive" context (b) contains the information that Julia is on her computer; this naturally implies that if she is writing an essay, she is doing so on her computer, so TYPE is implied and this is an *m*-positive context. Finally, the "infelicitous" context (c) is logically incompatible with the target utterance because in the context, Eliza doesn't see Julia in the library, but then in the video utterance she asserts that she saw Julia in the library.

We would expect participants shown this infelicitous context to rate the video as *not completely natural*, with or without a co-speech gesture.

As we mentioned above, another technique we used to create the infelicitous contexts was to swap the names of Eliza's interlocutor and the person under discussion in the video; this can be seen, for instance, in Scenarios 8, 9, and 10 in Appendix C. This was in an effort to keep the infelicitous context paragraphs as minimally different from the neutral contexts as possible, just as we strove to make neutral and positive contexts minimally different.

3.2 Results

Out of 90 participants, 2 participants' responses were excluded due to those participants failing at least one of the three attention checks. The results discussed below are for responses from the remaining 88 participants.

Table 5 reports the mean acceptance rates, standard deviations, and standard errors across trial types, and Figure 6 shows mean acceptance rates for each trial type with standard error bars.

Trends across the experimental conditions GESTURECUE and CONTEXT were as follows. On average, trials with gestures (trial types 4–6) were rejected slightly more often than trials without gestures (trial types 1–3) (with gesture: M = 0.62; without gesture:

Table 5: Experiment 2 trial types with conditions (GESTURECUE and CONTEXT) and mean acceptance rates.

Trial type	GESTCUE	CONTEXT	Mean	SD	SE	N
1	no	infel	0.403	0.492	0.032	233
2	no	neut	0.782	0.414	0.027	238
3	no	pos	0.815	0.389	0.025	233
4	yes	infel	0.449	0.498	0.033	234
5	yes	neut	0.696	0.461	0.030	230
6	yes	pos	0.725	0.447	0.029	240
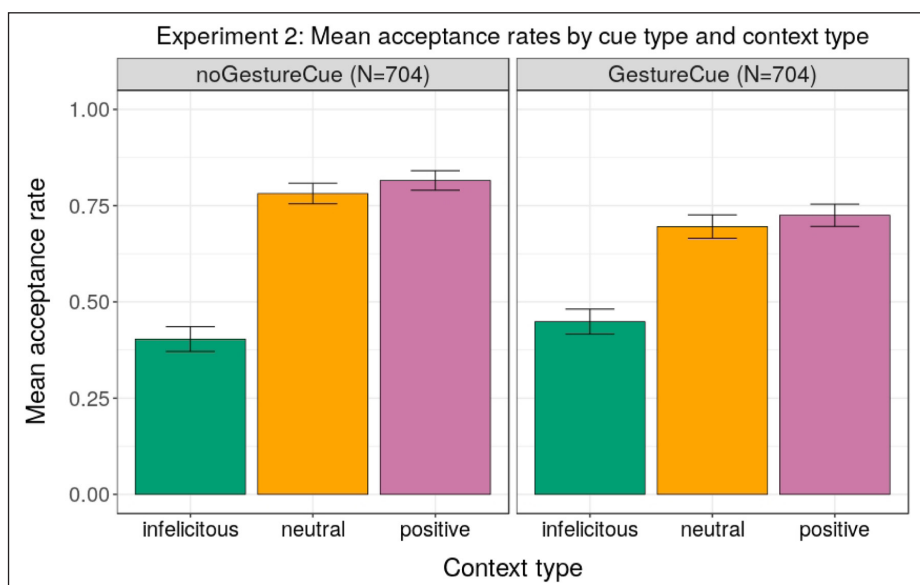
Figure 6: Experiment 2 mean acceptance rates, by experimental conditions GESTURECUE and CONTEXT.

M = 0.67). Trials shown with an infelicitous context (1 and 4) were rejected at a much higher rate than trials shown with either a neutral context (2 and 5) or a positive context (3 and 6) (infelicitous: M = 0.43; neutral: M = 0.74; positive: M = 0.77). Note that neutral and positive context trials were rejected at roughly the same rate.

Table 6 gives a breakdown of ratings by scenario. Means for scenarios across trials with infelicitous contexts ranged from 0.23 (Scenario 12) to 0.66 (Scenario 2), with SDs ranging from 0.43 to 0.51. Means for scenarios across trials with either neutral or positive contexts, by contrast, ranged from 0.67 (Scenario 16) to 0.93 (Scenario 5), with SDs ranging from 0.25 to 0.49.

Analyses of participants' judgment responses were conducted using the R programming language to build generalized linear mixed effects models using the function *glmer*. In the model with the most data variation coverage (determined by ANOVA testing), the independent factors in the model were GESTURECUE (no/yes) and CONTEXT (neutral/positive/infelicitous); RESPONSE (1/0) was the dependent variable, and scenario number and participant ID were coded as random effects.

Results of the model indicate that there was a significant main effect of CONTEXT: With "neutral" as the reference value, [CONTEXT = infel] was highly significant ($\beta = -1.796$, $z = -7.887$, $p < 0.001$), but [CONTEXT = pos] was not significant ($p > 0.1$). In a minimally different model with "infelicitous" as the CONTEXT reference value instead, both [CONTEXT = neut] and [CONTEXT = pos] were highly significant with ($\beta = 1.796$, $z = 7.887$, $p < 0.001$) and ($\beta = 2.034$, $z = 8.590$, $p < 0.001$), respectively. In other words, there was no significant difference between the neutral and positive context conditions, but an infelicitous context significantly decreased the acceptance rate compared to either neutral or positive contexts.

Unlike the results from Experiment 1, GESTURECUE was not a significant main effect ($p > 0.1$). We discuss possible reasons for this difference in Section 3.3. There were no significant interactions between factors.

Table 6: Experiment 2 acceptance rates by scenario, with infelicitous trials separated out from combined neutral and positive trials.

Scenario	M (infel)	M (neut/pos)	SD (infel)	SD (neut/pos)
1	.38	.71	.49	.46
2	.66	.69	.48	.46
3	.53	.79	.51	.41
4	.32	.68	.48	.47
5	.55	.93	.51	.25
6	.41	.80	.50	.41
7	.48	.74	.51	.44
8	.33	.79	.48	.41
9	.62	.80	.49	.41
10	.33	.76	.48	.43
11	.52	.85	.51	.36
12	.23	.79	.43	.41
13	.34	.61	.48	.49
14	.31	.78	.47	.42
15	.38	.68	.50	.47
16	.40	.67	.50	.47

We also analyzed the subset of participant responses for which the context presented was either "neutral" or "positive", thereby excluding the data from "infelicitous" CONTEXT trials. Again we used R and the the function *glmer* to build a generalized linear mixed effects model, with independent factors GESTURECUE and CONTEXT, dependent variable RESPONSE, and random effects of scenario number and participant ID. Results of the model indicate no significant main effect of either GESTURECUE or CONTEXT, nor a significant interaction between the two factors ($p > 0.1$). This supports the conclusion from the full model reported above that there was no significant difference in acceptance rates between trials with neutral contexts and those with positive contexts.

3.3 Discussion

Since the trials with infelicitous contexts were rated significantly lower than other trials, this follow-up study is encouraging, suggesting that participants were reading the context paragraphs and, just as importantly, taking them into account when rating the video utterances. Similarly to Experiment 1, in this follow-up study there was no significant difference in ratings between trials with neutral contexts and trials with positive contexts. Taken together, we can conclude more confidently that co-speech gestures are not sensitive to entailment/implication by the discourse context, and hence are [–SCF] in the Tonhauser et al. (2013) terminology.

The factor GESTURECUE was not a significant main effect in Experiment 2. We suspect that this is likely due to the overwhelming effect of context in Experiment 2 with the introduction of infelicitous trials, minimizing the variability available among felicitous trials. To investigate this further, we compared the means for the same trials across both experiments (the two trial types [SPEECHCUE = yes, GESTURECUE = no] and [SPEECHCUE = no, GESTURECUE = yes]), and found that the range of means for Experiment 1 is 0.60–0.77 (mean SD 0.46), compared to the more compressed range of means in Experiment 2 of 0.70–0.82 (mean SD 0.43). In other words, the same trials in Experiment 2 had higher and more compressed ratings for positive and neutral contexts than those trials did in Experiment 1, which we attribute to the addition of infelicitous contexts in Experiment 2. We therefore suspect that this compression of contexts of interest (felicitous trials) may have contributed to the lack of effect for GESTURECUE in Experiment 2 compared to Experiment 1.

In Section 3.1.4, we noted that many of the infelicitous contexts for this experiment were created by simply swapping the names of the addressee and the person being talked about in the neutral and positive contexts. One might wonder if participants were paying sufficient attention to these "name-swap"-type infelicitous contexts to detect the cause of infelicity, and if they were not, whether this artificially raised the ratings of those trials. To test this, we coded the scenarios for Experiment 2 according to whether their corresponding infelicitous contexts were made bad by swapping names ("name-swap") or by some other mechanism ("other"), such as introducing a contradiction between the context paragraph and the video utterance (e.g., the infelicitous context of Scenario 5, Appendix C). We constructed a generalized linear mixed effects model on the data from just those trials where the participant saw an infelicitous context (condition [CONTEXT = infel]), with RESPONSE as the dependent variable, type of infelicity (name-swap/other) and GESTURECUE as the independent variables, and subject ID and scenario as random effects. There was no main effect of type of infelicity on acceptance rates, nor a significant interaction effect between type of infelicity and GESTURECUE ($p > 0.1$ in both cases). We interpret these results as indicating that participants did not behave differently when presented with a "name-swap" infelicitous context than they did with other types of infelicitous contexts, and so the cause of infelicity did not have an overall effect on felicity judgments.

As in Experiment 1, we wanted to check whether the belief state of the addressee in the "dialogue" between Eliza and her interlocutor described in the context paragraphs influenced participants' judgments. We conducted another follow-up analysis on the Experiment 2 data, using the same classification of scenarios (yes/no) according to whether the addressee of the dialogue knows or could infer the content of the proposition in question based on the information provided by the "positive" context. We constructed two separate generalized linear mixed effect models, one on the "yes" subset of the data (where the addressee knows/can infer the semantic content of the gesture based on the positive context), and the other on the remaining "no" subset of the data, with RESPONSE as the dependent variable, GESTURECUE and CONTEXT as the independent variables, and subject ID and scenario number coded as random effects. Just as in the Experiment 1 analysis, neither model showed a significant effect of CONTEXT on RESPONSE ($p > 0.1$ in both cases). From this we once again conclude that the belief state of the addressee is not a significant factor in acceptability judgments of utterances with co-speech gestures based on prior discourse contexts.

Overall, Experiment 2 lends support to our experimental methodology, providing encouraging evidence that participants judged the overall effect of context plus video when providing their ratings, although the addition of the infelicitous contexts meant that more subtle differences between generally felicitous trials in some cases disappeared. Together, we suggest that Experiments 1 and 2 provide a more clear overall picture of the empirical landscape for co-speech gesture pragmatics.

4 Conclusions

4.1 Directions for future research

By directly controlling whether the content of co-speech gestures is duplicated in the preceding context and/or the same utterance, the two experiments we report in this paper show how the acceptability of co-speech gestures may be affected by linguistic context. We hope that this can be a first step for future studies, and, in some cases, provide a foundation for interpreting results found in the few already existing studies on the semantics/pragmatics of co-speech gesture.

Through the notion of triviality, we connect to the existing literature on the way that gestures contribute content, which has been the focus of previous discussion of co-speech gestures as either "supplemental" (Ebert & Ebert 2014; 2016) or as "cosuppositional" (Schlenker 2018). Under a supplemental analysis, it is quite surprising that speech cues aid the acceptability of gestures, given that supplements are typically less acceptable if they are trivial. On the cosuppositional side, we have shown that co-speech gestures are not "hard" presupposition triggers since they need not be entailed by preceding context; however, many presuppositions are known to be easily accommodated, which could account for their acceptable nature in our study. Under this view, it remains a question why a positive preceding context doesn't also provide the same kind of content support (and our second experiment suggests that it is not due to participants' ignoring written context). Altogether, given the *improvement* of gestures with matching speech cues, our data are much harder to reconcile with the supplemental theory.

One further remaining question raised by these studies has to do with the choice of gesture targets in the development of the stimuli, discussed above in the Methods section (2.1.3) of Experiment 1. Recall that we chose semantically underspecified predicates so that the corresponding co-speech gestures would non-trivially contribute to the truth conditions of the utterances. Some underspecified predicates seem to have a preferred "default" manner or adjectival property that is assumed in the absence of further modification (either by speech cue or gesture cue). Ebert & Ebert (2016) look at some

examples of these default interpretations for NPs in German, referring to the phenomena as the "typicality of a gesture for an NP concept". A particularly interesting example they describe involves the NP *Fenster* 'window' and the two different shape properties of either square or circular; they assume the square option to be the more typically expected shape for windows. For an example from our experimental stimuli, consider the VP *writing an essay*:

(30)   Scenario 5 (Appendix A)
        I saw Julia over in the library [writing an essay]_TYPE – it looks like she's a
        little preoccupied right now.

In this scenario we chose the gesture TYPE, indicating that the writing was done on a computer as opposed to by hand with pen and paper. In the current decade, it would be a very natural inference to make that *writing an essay* means writing (typing) on a computer; anyone visiting a college campus these days will see ample evidence of this preferred mode of writing essays. Returning to the experiment, in Scenario 5 trials such as (30) we chose the gesture option with the "default" or more expected interpretation of TYPE instead of a more uncommon co-speech gesture indicating writing with a pen and paper, call it WRITE-WITH-PEN. An important question is whether this choice of gesture has a significant effect on utterance ratings and whether it interacts in a significant way with the presence of a speech cue in an utterance. The supplement analysis of co-speech gesture, although generally not supported by our findings, would predict that less trivial, less expected gestures (given speech cues) would be improved, and this does match our intuitions about examples like (2) from Goldin-Meadow & Brentari (2017), repeated below as (31), which we find especially natural/felicitous:

(31)   I [ran up the stairs]_SPIRAL-UP.

Given that the prototypical staircase is not a spiral staircase, the gesture content is unexpected and hence informative. An experiment could be designed with gestures varying along the dimension of unexpectedness, given context and speech cue.

4.2 Summary

The primary goal of these experiments was to diagnose the behavior of co-speech gestures in contexts that do and do not entail/imply their semantic contents, thereby gaining a better understanding of when and how co-speech gestures can be felicitously used in conversation. We implemented this question formally using the strong contextual felicity diagnostic proposed in Tonhauser et al. (2013), and by varying whether similar information was contained in accompanying speech cues. The results of our experiments show that co-speech gestures do not need to be entailed/implied by their preceding discourse context; hence co-speech gestures are [–SCF] and cannot be considered "hard" (i.e., unaccommodatable) presupposition triggers. We also saw that *speech cues*, or speech expressions that (approximately) duplicate the semantic content of a co-speech gesture, had a significant interaction with the presence or absence of a gesture in the trials. We take this to be an indication that there are other restrictions on the felicitous use of co-speech gestures that we do not yet know about and that involve when and how the gestures can contribute "extra" semantic content to the utterance meaning. We speculate that this may vary depending on the type of content conveyed in the gestures, for example whether they represent size-and-shape or manner information, and whether they modify nouns or verbs, among other potentially relevant dimensions. We hope that the data

on co-speech gesture felicity judgments gathered through these experiments will pave the way for future research on co-speech gestures that addresses these larger theoretical questions in interesting and fruitful ways.

Abbreviations

SCF = strong contextual felicity

Additional Files

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

- **Appendix A.** Stimuli for Experiment 1. DOI: https://doi.org/10.5334/gjgl.438.s1
- **Appendix B.** Gesture screencaps. DOI: https://doi.org/10.5334/gjgl.438.s1
- **Appendix C.** Stimuli for Experiment 2. DOI: https://doi.org/10.5334/gjgl.438.s1

Acknowledgements

Competing Interests

The authors have no competing interests to declare.

References

Abner, Natasha, Kensy Cooperrider & Susan Goldin-Meadow. 2015. Gesture for linguists: A handy primer. *Language and Linguistics Compass* 9(11). 437–449. DOI: https://doi.org/10.1111/lnc3.12168

Abrusán, Márta. 2016. Presupposition cancellation: Explaining the 'softhard' trigger distinction. *Natural Language Semantics* 24(2). 165–202. DOI: https://doi.org/10.1007/s11050-016-9122-7

Baayen, R. Harald, Douglas J. Davidson & Douglas M. Bates. 2008. Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59(4). 390–412. DOI: https://doi.org/10.1016/j.jml.2007.12.005

Bates, Douglas M., 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

Chierchia, Gennaro & Sally McConnell-Ginet. 2000. *Meaning and grammar: An introduction to semantics*. Cambridge, MA: MIT Press.

Ebert, Cornelia & Christian Ebert. 2014. Gestures, demonstratives, and the attributive/referential distinction. Slides from a talk given at *Semantics and Philosophy in Europe (SPE 7)*, ZAS, Berlin, June 2014. Available at: https://semanticsarchive.net/Archive/GJjYzkwN.

Ebert, Cornelia & Christian Ebert. 2016. The semantic behavior of co-speech gestures and their role in demonstrative reference. Slides from an LSCP LANGUAGE seminar talk at the Institut Jean-Nicod, Département d'Études Cognitives, École Normale Supérieure, Paris, Dec. 2016. Available at: http://www.cow-electric.com/neli/talks/CE-Paris-2016.pdf.

Esipova, Maria. 2018. Focus on what's not at issue: Gestures, presuppositions, appositives under contrastive focus. *Proceedings of Sinn und Bedeutung 22 (SuB 22)* (to appear). Available at: https://ling.auf.net/lingbuzz/003892.

Goldin-Meadow, Susan & Diane Brentari. 2017. Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences* 40. e46. DOI: https://doi.org/10.1017/S0140525X15001247

Kendon, Adam. 1980. Gesticulation and speech: Two aspects of the process of utterance. In Mary Ritchie Key (ed.), *The relationship of verbal and nonverbal communication*, 207–227. The Hague: Mouton.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511807572

Lascarides, Alex & Matthew Stone. 2009. A formal semantic analysis of gesture. *Journal of Semantics* 26. 393–449. DOI: https://doi.org/10.1093/jos/ffp004

McNeill, David. 1985. So you think gestures are nonverbal? *Psychological Review* 92(3). 350–371. DOI: https://doi.org/10.1037/0033-295X.92.3.350

McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. Chicago, IL: University of Chicago Press.

McNeill, David. 2005. *Gesture and thought*. Chicago, IL: University of Chicago Press. DOI: https://doi.org/10.7208/chicago/9780226514642.001.0001

Potts, Christopher. 2005. *The logic of conventional implicatures*. Oxford: Oxford University Press.

R Core Team. 2016. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. https://www.Rproject.org.

Schlenker, Philippe. 2018. Gesture projection and cosuppositions. *Linguistics and Philosophy* 41. 295–365. DOI: https://doi.org/10.1007/s10988-017-9225-8

Tieu, Lyn, Robert Pasternak, Philippe Schlenker & Emmanuel Chemla. 2017. Co-speech gesture projection: Evidence from truth-value judgment and picture selection tasks. *Glossa: a journal of general linguistics* 2(1). 1–27. DOI: https://doi.org/10.5334/gjgl.334

Tonhauser, Judith, David Beaver, Craige Roberts & Mandy Simons. 2013. Toward a taxonomy of projective content. *Language* 89(1). 66–109. DOI: https://doi.org/10.1353/lan.2013.0001