# Testing contrastive inferences from suprasegmental features using offline measures

Anna Alsop, Elaine Stranahan & Kathryn Davidson[*]

**Abstract.** Speakers add modifiers to the extent that they are informative (Grice 1975); studies using the visual world eye-tracking paradigm find that the use of pre-nominal modifiers (*short, big*) leads listeners to infer the existence of similar objects differing along that same scale (Grodner & Sedivy 2011; Sedivy et al. 1999). In this study, we probe these contrastive inferences using an offline questionnaire, paired with audio/video stimuli to ask whether similar inferences extend to two types of supraseg-mental features: prosodic focus and depictive co-speech gestures. Our results suggest that the presence of a scalar adjective robustly leads to contrastive inferences in this offline forced choice paradigm, and that the robustness of the lexical pattern persists even when prosodic focus would indicate otherwise. Prosodic focus does, however, appear to modulate the contrastive effect of a given pre-nominal modifier. We find that the same pragmatic process fails to extend to depictive co-speech gestures, supporting a semantic analysis of these gestures as generally not-at-issue contributions.

**Keywords.** experimental pragmatics; contrastive inference; focus; co-speech gesture

**1. Introduction.** When engaged in conversation, speakers and listeners follow certain fundamental conversational expectations to communicate effectively and efficiently (Grice 1975). We focus in this paper on expectations governing the tradeoff between informativity and brevity: speakers should be highly informative, but also as succinct as possible, and these two pressures often conflict. In particular, we will use this tradeoff as a starting point to examine the pragmatics of noun modification, asking when a speaker may choose to forgo brevity to include a modifier in an utterance, and how two suprasegmental features (prosody, and co-speech gesture) interact with this calculation. In doing so, we will also present a new offline experimental paradigm to investigate pragmatic inferences involved in modifier interpretation.

**2. Background: Contrastive inference.** One conversational implicature, well-attested in experimental literature, that arises from noun modification is known as *contrastive inference* (Sedivy et al. 1999). Contrastive inferences occur when a speaker uses a pre-nominal modifier to restrict the meaning of a noun phrase, as in sentence (1-a).

(1)    a.    Give me the tall glass.
       b.    Give me the glass.             (adapted from Sedivy et al. 1999)

Due to the pragmatic expectation that a speaker will only provide as much information as is required, the modifier *tall* must be maximally informative in the discourse context, else the speaker would have chosen (1-b). Moreover, the speaker has a strong preference for brevity, so the modifier must make a crucial contribution to the utterance in order to warrant its inclusion. The listener then infers that *tall* serves to disambiguate the goal referent from a similar, not tall object.

More generally, in order for the scalar adjective to be informative, there must be at least one other salient object of the same type but differing along the given scale (Grodner & Sedivy 2011; Levinson 2000; Sedivy 2003; Sedivy et al. 1999). In an early study by Sedivy et al. (1999) of

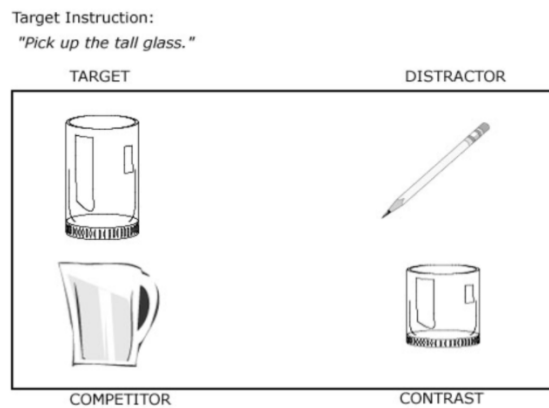Target Instruction:
"Pick up the tall glass."



Figure 1: Visual display from Grodner & Sedivy (2011)

contrastive inference computation, real time listener comprehension was monitored via the visual world eye-tracking paradigm. Sedivy et al. presented participants with a two-by-two visual display (see Fig. 1), and monitored eye movements across the display as participants listened to a set of recorded instructions (Pick up the ADJ NOUN). Each display contained a minimal pair of two objects of the same type, differing with respect to a certain property: the target object (correctly described by the instruction), and the contrast object. The competitor object was of a different object type but shared the target property, and the fourth item was a distractor.

A listener processing the literal meaning of the instruction would only be able to fully disambiguate the referent after hearing the complete utterance, since the target and competitor objects share the same property. However, when participants listened to the instruction up until "Pick up the ADJ- ", their eyes fixated on the target object early, presumably inferring that use of the adjective implied a contrast set. This early referent disambiguation has been replicated in subsequent eye tracking studies, and offers strong support for a contrastive interpretation of pre-nominal modifiers (Grodner & Sedivy 2011; Pogue et al. 2015; Sedivy 2003). That said, not all modifiers elicit contrastive inference: for example, in sentence (2-a), *yellow* is a generally uninformative descriptor of a banana because yellow is already a default property of bananas, while *purple* is a highly informative descriptor of a banana because so few bananas are in fact purple (see (2-b))

(2)  a.  Give me the yellow banana.
     b.  Give me the purple banana.                            (adapted from Sedivy 2003)

The first have been shown to trigger a more exaggerated contrastive inference effect because the modifier is so rarely used to describe the object in isolation (Sedivy 2003), and when nouns are combined with unexpected and atypical modifiers as in the case of *purple banana*, no contrastive inference is observed. With these types of utterances, the modifier is used to note an unusual instance of a kind, rather than to signal contrast. Further studies have revealed additional criteria for pragmatic inference computation, especially that contrastive inference is only generated when

the speaker is assumed to be cooperative and reliable (Grodner & Sedivy 2011; Pogue et al. 2015).

Any viable pragmatic framework should account for the basic contrastive inference, inferences about object atypicality, and the cancelability of such inferences. In this paper, we have three aims related to studying contrastive inference. First, we propose and implement an offline forced choice task. This novel experimental paradigm differs from previous online paradigms in multi-ple respects: it has the potential to capture listener post-hoc reasoning about an utterance, which may reveal further listener inferences that are more cognitively demanding and take more time to compute, and it asks participants to reconstruct a world state from an utterance, a fundamentally different task from previous experiments in which participants simply arrived at a single referent.

Using this offline task, we investigate a question raised by Sedivy (2003) concerning the relationship between focus and pre-nominal modifiers, which appear to generate a similar contrastive effect. Focus, marked prosodically by a pitch-accent and change in rhythm, triggers the generation of a set of alternatives from which the normal semantic value is taken (Rooth 1992). This conveys contrast between the set of alternatives and foregrounded semantic value, just as a pre-nominal modifier implies a contrast set of objects differing by the modifier property. According to preva-lent theories in focus semantics, we would expect that focus placement to play a role in contrastive inference such that contrastive inference should only and especially arise when a pre-nominal mod-ifier is focused, because this would generate the desired set of alternatives (Büring 2003; Roberts 2012; Rooth 1992). Focusing the head noun would instead result in a focus semantic value of ob-jects of different kinds, but sharing the modifier property–a different sort of contrastive inference. In our first experiment, we use the proposed forced choice task to test the effect of prosodic focus on the pragmatic interpretation of modifiers, analyzing the results to better understand the role of information structure in inference computation.

Finally, in a second experiment, we expand on our study of suprasegmental features to ask whether other kinds of modifiers lead to a contrastive interpretation, probing the effect of infor-mation contributed via co-speech gesture. Depictive co-speech gestures occur at the same time as speech, and add meaning through clear visual and spatial representation integrated with linguistic content (McNeill 1992; Kendon 2004; Goldin-Meadow & Brentari 2015). In recent work in formal semantics/pragmatics, Ebert & Ebert (2014) and Schlenker (2018) observe that co-speech gestures typically are not part of the actively asserted content of an utterance. We hypothesize that this backgrounded informational status, as opposed to the at-issue status of lexical adjectives, may result in a different status for contrastive inferences based on co-speech gestures, and test this using the same offline methodology.

## 3. Methodology: An offline experimental paradigm for contrastive inference.

3.1 MOTIVATION. The visual world eye-tracking paradigm is popular in psycholinguistics as a methodology effective for capturing incremental language processing. When presented with spo-ken instructions, participants' eye movements to certain objects have been shown to be time-locked to the occurrence of words that refer to them (Sedivy et al. 1999; Tanenhaus et al. 1995). These eye movements are thus a robust proxy for the linguistic processing occurring at any given point in spoken utterance interpretation, and its interaction with a visually presented context. This online eye-tracking paradigm has been shown to be a sensitive measure for spoken utterance processing, providing the first empirical evidence for contrastive inference. However, offline measures may

also offer several advantages in other ways. First, an offline paradigm may provide more data on pragmatic inferences that take more time and more extensive pragmatic reasoning to compute, while an online paradigm more effectively captures quick inferences. If the goal is to capture a listener's final interpretation upon the weighing of several incrementally computed inferences at the end of an utterance, an offline measure may be more efficient. Second, offline measures more naturally allow for the study of more visual language, such as gesture and sign languages. While audio instructions leave subjects' eyes free to move across a display of objects, video stimuli would be difficult to integrate into the paradigm because it would divert participant eye movement away from the objects. Third, if we are interested in collecting a large body of cross-linguistic data, an offline paradigm would lessen the burden of participant recruitment, opening up avenues such as social media and crowd-sourcing platforms (e.g. Amazon Mechanical Turk). We decided then to create a fundamentally different task to test contrastive inference; whereas previous eye-tracking studies simply instructed participants to choose an object from a display, this novel task instructs participants to infer the objects in the world state based on an utterance.

3.2 OFFLINE FORCED CHOICE TASK. In our offline experimental paradigm, participants are presented with a scenario in which two characters are engaged in dialogue. One character presents two objects at a time to the second character, who then indicates which of the two he/she wants. The second character's spoken requests are all of the form, "Give me the ADJ NOUN".

It is clear from a photo presented in the instructions that the objects are visible to both characters, but not to participants, who are only given the second character's requests. Participants are then asked to reason about which set of objects the characters saw. If the task elicits contrastive inferences, participants should choose the pair of similar items when an extra modifier is present (for thick book, the left display in Figure 2), and choose the pair of dissimilar items when the modifier is absent (for book, the right display in 2). We will refer to the display with similar items as the "adjective-contrast" display, because the contrast exists in the adjective property. The other display will be referred to as the "noun-contrast" display, because the two objects differ in object type, encapsulated by the noun in each sentence.

This offline forced choice task can be straightforwardly implemented in a questionnaire. We used the Qualtrics Survey Creation Tool as the platform for both of our experiments, which allowed for random assignment of participants, randomization of trial order, and counterbalancing.



Figure 2: Sample forced choice task with the request, "Give me the thick book."

For our experimental stimuli, we chose everyday household objects, many of which over-lapped with the experimental stimuli in previous eye-tracking studies (Grodner & Sedivy 2011;

Sedivy 2003; Sedivy et al. 1999). We used images pulled from open-source image databases, which were high-quality, realistic illustrations of the stimulus objects. We also made the decision to use scalar and dimensional adjectives, such as *big*, *short*, and *round*. There were several practical considerations taken into account for this choice, including that our experiments were computer-based so some percepts are easier than others to convey on screen, and scalar adjectives are conveyable via depictive co-speech gesture, which we tested in our second experiment.

## 4. Experiment 1: Prosodic Focus.

4.1 FOCUS AND CONTRASTIVE INFERENCE. In her discussion of contrastive inference, Sedivy (2003) notes that the contrastive effect of pre-nominal modifiers resembles that of focus marking. Semantic focus $[\ ]_F$ is marked by a pitch-accent and change in rhythm and triggers the generation of a set of alternatives from which the normal semantic value is taken (Rooth 1992). This conveys contrast between the set of alternatives and foregrounded semantic value, just as a pre-nominal modifier implies a contrast set of objects differing by the modifier property.

Aside from generating a contrast set of alternatives, focus-marking has been shown to mark agreement in the information structure of the discourse (Roberts 2012). At each point in discourse, there is one or more salient Questions Under Discussion (QUDs), which speakers are expected to answer. The presupposition conveyed by focus in an assertion is that the assertion is relevant, at least partially answering a salient QUD. Following this line of reasoning, prosodic focus is only felicitous if the focus semantic value of an assertion matches the alternatives of a QUD (Roberts 2012).

To illustrate, (3-b) is a felicitous assertion when (3-a) is the salient QUD because the alternatives of both (3-a) and (3-b) are the set of possible glasses. The same utterance ((3-b) is written again as (4-b) for ease of comparison) is infelicitous in the context of QUD (4-a); there is a mismatch between the QUD alternative set (all possible objects) and the focus semantic value of the assertion (all possible glasses). This mismatch is resolved if the intonational focus in the assertion is removed, as in (4-c).

(3)    a.    Which glass should I give you?
       b.    Give me the $[\text{tall}]_F$ glass.

(4)    a.    Which object should I give you?
       b.    #Give me the $[\text{tall}]_F$ glass.
       c.    Give me the tall glass.

The role of focus-marking in encoding discourse structure should then lead it to interact with the pragmatic interpretation of the stimulus sentences in previous contrastive inference studies. Eberhard et al. (1995) specifically tested listener processing of instructions in which the pre-nominal modifier received prosodic focus. When the eye-tracker data was compared against results of control trials in which the modifier received neutral stress, there was a significant increase in rate of fixation on the target object in the display.

Sedivy et al. (1999) conducted a later follow-up study to replicate these results, recording instructions focusing the pre-nominal modifier. The control trials again contained stimulus sentences with neutral intonation. However, when these results were compared, no main effect of word stress on contrast effect was observed. Sedivy et al. point out that their stimulus sentences

did not always contain adjectival modifiers, while all of Eberhard et al.s did; subjects in the Sedivy et al. study may then have taken pre-nominal modifiers to be more informative, causing the difference in results. Though this particular difference in experimental design no doubt influenced the resulting data, further inquiry is necessary to better understand the effect of focus in listener contrastive inference computation.

Given the relationship between focus and alternatives in pragmatic theory, the manipulation of focus placement should lead (5) and (6) to have different focus semantic values, and consequently different interpretations. We would predict (5) to contrast a tall glass against a set of alternative glasses, and (6) to contrast a tall glass against a set of tall, non-glass objects.

(5)    Give me the $[\text{tall}]_F$ glass.

(6)    Give me the tall $[\text{glass}]_F$.

The existing contrastive inference literature would suggest that the mere presence of the added scalar adjective always leads to a contrast set of alternative glasses, contradicting the semantics-informed prediction for the noun-focused sentence. In this experiment, we ask which of these two predictions more accurately describes listener interpretations of sentences like (5) and (6).

4.2 METHODS.

4.2.1 Participants

40 adult native speakers of English, all located in the United States, were recruited via Amazon Mechanical Turk. 3 participants failed catch trials, and were excluded from final data analysis. Each participant was compensated with a payment of $1.50 through Amazon payments.

4.2.2 Procedure

We created a computer-based questionnaire using the Qualtrics Survey Creation Tool and recruited participants through Amazon Mechanical Turk. In the instructions, the participant was familiarized with the forced choice task with an example display containing shapes. The example task has a correct answer but participants were informed that not all questions would have a "right" answer; in these trials, they should choose answers that "makes the most sense" to them. This caveat was included to encourage task completion based on contrastive inference when the literal utterance interpretation does not provide sufficient information.

Next, the participant was presented with a series of trials, in which Annie, the speaker in the scenario, uttered a spoken request in the form "Give me the ADJ NOUN". The speaker identifying as Annie in the audio recordings was a female native English speaker in her early twenties. In each trial, participants were instructed to click to listen to the audio clip (embedded using SoundCloud), which they could replay if needed. They were then forced to choose the pair of objects they believed were on the table between Annie and Lana.

Participants completed 20 critical trials, in which utterances were paired with a forced choice task. In all of the critical trials, both display choices contained the same referent requested by Annie, but one display contained a contrast set of two objects of the same type, and the other, a contrast set of two objects of different types. Along with the critical trials, participants were presented with 2 catch trials intended to identify participants who were not playing the audio clips,

or not paying sufficient attention to the task. The catch trials were visually indistinguishable from the critical trials, except only one of the displays contained the correct object requested by Annie. Thus, each catch trial had a clear correct answer; we excluded data from the 3 participants who did not answer both catch trials correctly. Overall, we created 22 pairs of stimulus objects differing on some scale involving size or shape (See Figure 3 for example image pairs).

4.2.3 Design

As outlined in the procedure above, each participant completed 22 trials, consisting of 20 experimental trials and 2 catch trials presented in a randomized order. Experimental trials fell into one of two prosodic conditions: noun-focused or adjective-focused. We used a within-subjects design, so that each participant saw an equal number of experimental trials in each prosodic condition.

Each stimulus item had two associated modifiers (e.g. tall/short glass), which were manipulated across trials so that each participant saw each item only described by one adjective. This resulted in four stimulus lists, ensuring that the four conditions of each stimulus item were seen by an approximately equal number of participants (see Table 1). The stimulus displays and individual objects within each display were also counterbalanced for relative position.

The Qualtrics questionnaire randomly assigned each participant to one of the four stimulus lists, and the 22 trials were presented in a randomized order.
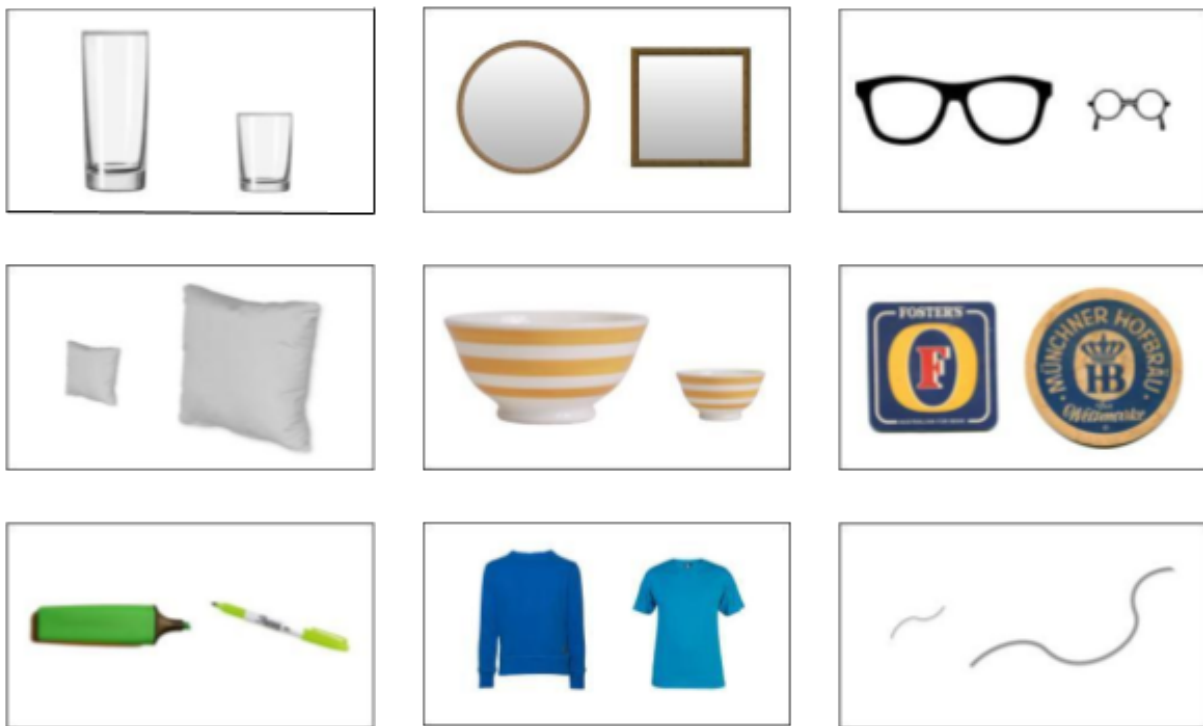


Figure 3: Example stimulus displays from Experiment 1 (top row, left to right: tall/short glass, round/square mirror, big/small glasses; middle row: small/big pillow, big/small bowl, square/round coaster; bottom row: thick/thin marker, long-sleeved/short-sleeved shirt, short/long string)

| Condition | Focus | Adjective | Utterance |
|:---:|:---:|:---:|:---:|
| 1 | Noun | Thick | Thick [book]$_F$ |
| 2 | Noun | Thin | Thin [book]$_F$ |
| 3 | Adjective | Thick | [Thick]$_F$ book |
| 4 | Adjective | Thin | [Thin]$_F$ book |

Table 1: The four experimental conditions of stimulus item Book

4.3 RESULTS. Excluding the three participants who failed the catch trials, we analyzed the data from the remaining 37 participants. We first analyzed overall participant behavior, examining for each condition the proportion of participants who chose the display containing the contrast set of two objects of the same type (e.g. tall glass and short glass). The proportion of participants choosing this adjective-contrast display was over chance for both focus conditions (see Table 2).

We performed a logistic mixed effects regression using the glmer function in R (version 3.4.0), with the prosodic focus as the independent variable. The dependent variable was binary, and indicated whether a participant selected the display containing two objects of the same type; in other words, whether a participant reported contrastive inference. Item, adjective presented with an item, and participant ID were included as random effects.

There was a significant main effect of noun focus on contrastive inference elicitation; focusing the noun rather than the adjective decreased participant reporting of contrastive inference ($p < 0.01$). Most items received similar responses, in which the contrastive inference was overall above chance for both focus conditions. Across items, adjective-focused trials generally generated more robust contrastive inferences than noun-focused trials.

Finally, an interesting pattern arose when examining the responses of each individual participant. The majority of participants (60%) reported contrastive inference most (>75%) of the time, as predicted by the contrastive inference literature. The next largest group (29%) reported the opposite contrastive inference, selecting the noun-contrast display most of the time (selecting the adjective-contrast <30% of the time). The smallest group (11%) behaved as predicted by focus theory, selecting the adjective-contrast display in adjective-focused trials.

4.4 DISCUSSION. The results of Experiment 1 support the robustness of contrastive inference reported in previous eye-tracking studies. Although contrastive inference was reported more frequently in adjective-focused trials, both focus conditions yielded an above-chance rate of adjective-contrast inference reports. Even when pitted against the effect of noun-focused prosody (which would otherwise be expected to lead to bias for two different object types, or noun-contrast), the

|  | Proportion of adjective-contrast responses |
|:---|:---:|
| Adjective-focused | 71% |
| Noun-focused | 61% |

Table 2: Proportion of adjective-contrast responses (%) by focus condition

pragmatic effect of the added adjective was sufficient to result in majority adjective-contrast display choices.

Experiment 1 also served as a test of our offline experimental paradigm. The individual participant response data was favorable, suggesting that the majority of participants were aware of their own contrastive inferences, and able to take these inferences into account when reasoning about discourse and world states. Subsequent experiments can also include trials without adjectives ("Give me the NOUN"), allowing for direct analysis of the effect of an added adjective in the forced choice task.

## 5. Experiment 2: Gestural modification.

5.1 SEMANTICS OF DEPICTIVE CO-SPEECH GESTURE. Gesture has long been the focus of work in social anthropology and cognitive psychology; only recently has it also begun to garner interest within linguistics, surprising since gesture is known to integrate with verbal communication in a meaningful and interpretable manner, combining with speech to produce a composite utterance (Kendon 2004; McNeill 1992). Gestures often occur in the same time slot as speech, and may carry a wide variety of meanings to supplement verbal content. In Experiment 2, we apply the offline forced choice task to gestural modification to shed light on its semantic-pragmatic properties.

The focus of this study will be depictive co-speech gesture, a subcategory of gesture that co-occurs with speech and adds meaning through visual and spatial depiction. Like lexical adjectives, these co-speech gestures may be used to modify nouns. However, unlike lexical adjectives, they are non-verbal, depictive, and occur in the same time slot as the target noun. These differences seem to lead to important differences in semantic integration. For example, Ebert & Ebert (2014) and Schlenker (2018) have both proposed formal semantic analyses for depictive co-speech gesture that accounts for the observation that they are typically not at-issue modifiers. They differ in implementation: Ebert & Ebert describe the semantic content of co-speech gesture as equivalent to that of a supplement, similar to an appositive phrase (e.g. "The pet, *a small black dog*, drank some water."). Schlenker, on the other hand, analyzes co-speech gesture as a weak trigger for a type of presupposition he dubs 'cosupposition. Co-speech gestures, Schlenker argues, convey information assumed to be true in the context of the speech they accompany. Thus, modifying information contributed by gesture is not part of the asserted content, but still serves to specify the meaning of the expression it modifies.

Example 12 illustrates the two differing analyses of sentence (7-a), in which a gesture roughly corresponding to the adjective "short" is used to modify "glass". (7-b) expresses the gestural content in a supplemental phrase (Ebert & Ebert 2014), and (7-c) includes the gestural content in a cosupposition (Schlenker 2018).
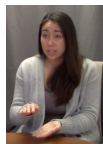


(7)　　a.　　Molly gave me a [SHORT][1] glass.
　　　　b.　　Molly gave me a glass, which by the way is short.

---

[1]Co-speech gestures will be represented by this bracket notation. Gestures should be interpreted as co-occurring with the spoken word that immediately follows.

      c.     Molly gave me a glass. ⤳ If x is what Molly gave me, and x is a glass, then it is short.

Clearly while these analyses differ in approach, both agree that co-speech gesture is not part of the asserted content, in contrast to the at-issue pre-nominal modifiers studied in Experiment 1.

     One way to see the difference is the contrast between the ability of co-speech gestures and lexical adjectives to be targeted by negation. For example, the adjectival modifier "big" in (8-a), and can be directly negated: in (8-b) it has to be true that any cat Molly gave the speaker is not big. This sentence is still true if Molly gave the speaker a small cat. On the other hand, the negated sentence (9-b) is not equivalent to (8-b), because the gestured modifier [BIG] appears to project out of negation. Only the assertion without the gestural modifier ("Molly didn't give me a cat") has been negated, so sentence (9-b) would be false even if Molly gave the speaker a small cat. Co-speech gestures have been shown to generally escape sentential negation (and other operators) in this way, motivating the categorization of gesture as not-at-issue.

(8)    a.     Molly gave me a big cat.
        b.     Molly didn't give me a big cat.

(9)    a.     Molly gave me a [BIG] cat.
        b.     Molly didn't give me a [BIG] cat.

In considering the pragmatic interpretation of gesturally conveyed information, then, two conflicting predictions arise. In (7-a), the specified property of height is gestured rather than verbally expressed. If similar maxims of conversation readily applies to the gestural domain, the addition of a modifying gesture [SHORT] would suggest that the speaker is contributing information crucial for referent disambiguation, resulting in contrastive inference. However, gesture fundamentally differs from speech in many respects, and may very well pass through a separate framework of interpretation. If gesturally-conveyed information does not behave according to Gricean pragmatics, then we would expect participants to draw contrastive inferences less reliably when presented with gestural modifiers.

     Underlying the resolution of these conflicting predictions is one final fundamental question: is gesture even intended to be an act of communication to begin with? Cooperrider (2018) outlines two types of gestures: those that are merely cognitive byproducts of language production, and those that are purposeful acts of communication. Co-speech gesture, Cooperrider argues, falls into the former category. McNeill (1992) also famously characterized gestures as "unwitting accompaniments" to speech, acts that are not deliberately produced or closely monitored by the speaker. If this is indeed the case for co-speech gestures, then they may be unlikely to set off the chain of pragmatic reasoning necessary for contrastive inference.

5.2 STIMULUS CREATION AND NORMING. To test whether co-speech gesture elicits contrastive inference similarly to lexical adjectives, we created a set of gestural modifiers corresponding roughly to the meanings of adjectives used in Experiment 1. However, we recognize that due to fundamental differences in the modalities of speech and physical gesture, these gestural and verbal modifier pairs are not likely to communicate exactly the same meaning, and so more than similarity to speech, our goal was simply to use gestural modifiers that could help a listener disambiguate each pair of similar objects.

Each of the 24 stimulus items (e.g. teddy bear) had two associated adjectives (e.g. small/big), resulting in a total of 48 corresponding gestural modifiers. Videos were filmed for each of the 48 gestural modifiers, in which the same speaker from Experiment 1 requested an item from the listener. The speaker was facing the camera in all videos, and the listener's back was to the camera, with the back of her head in frame. The table between the two interlocutors was partially visible as well (See Fig. 5 for a screen shot). All utterances were of the form "Give me the (ADJ) NOUN", and gestural modifiers were articulated in the same time slot as the lexical adjective.

In order to determine the acceptability of our gestural modifiers, we conducted a norming study (n = 80) with the 48 video and image pairs, presented in a Qualtrics online survey. Participants were all native speakers of English located in the United States, recruited on Amazon Mechanical Turk. The experimental scenario presented in the beginning of the survey was consistent with that of Experiment 1, in which speaker Annie is requesting items at a yard sale. In each trial of the study, participants were presented with an image of a single object, with the accompanying text, "Annie wants the object below." Below the image, participants were instructed to play the video of Annie asking for the object, and asked to report the naturalness of Annie's request on a 5-point Likert Scale.

Items received high ratings overall in the norming study, with a mean rating of 4.0. For the final stimuli, we selected the 20 video and image pairs that received the highest ratings.

## 5.3 METHODS.

### 5.3.1 Participants

80 adult self-reported native speakers of English with locations in the United States were recruited via Amazon Mechanical Turk. Participants included in Experiment 1 were disqualified from Ex-periment 2. One participant failed catch trials, and was excluded from the final data. Each partici-pant was compensated with a payment of $1.00 through Amazon payments.

### 5.3.2 Procedure

As in Experiment 1, Experiment 2 was conducted using a Qualtrics online questionnaire. Participants were provided with a consent form and instructions at the start of the questionnaire. After participants gave consent, they were presented with task-specific instructions modeled after Experiment 1. The instructions outlined an experimental scenario in which Annie and Lana are at a
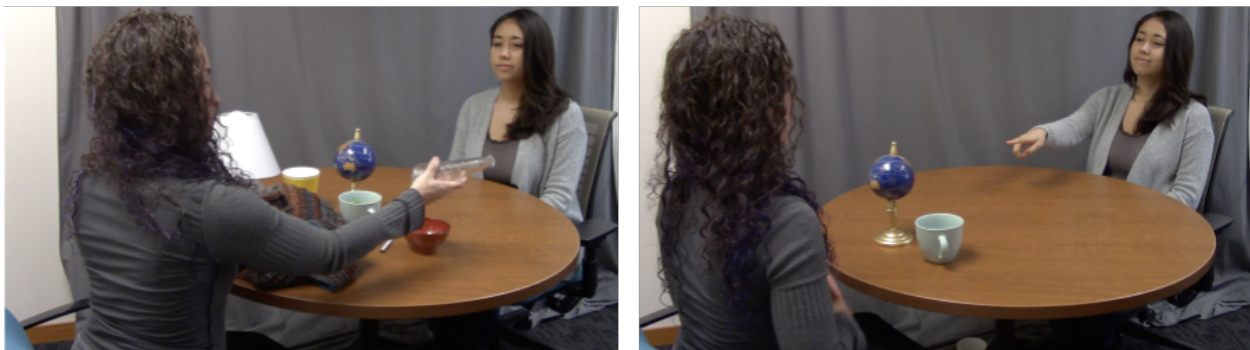


Figure 4: Images of experimental scenario, presented to participants in the instructions

yard sale, with Lana placing two items at a time on the table. The instructions also provided still images of Annie and Lana at the table for context (See Fig 4). As in Experiment 1, participants were informed that not all questions would have a right answer, and that they should choose what "makes most sense" to them.

After reading the instructions, each participant completed 20 trials, in which they were instructed to watch a video of Annie requesting one of the two items. Videos could be replayed as needed, and all contained utterances of the form, "Give me the (ADJ) NOUN", with or without co-speech gesture (See Fig. 5 for an example trial).

### 5.3.3 Design

Experiment 2 manipulated two variables of interest: gestural modification and linguistic modification. The combination of these two variables resulted in 4 total conditions, defined by the presence of gesture and lexical adjective. Video descriptions, one for each of the conditions ($\pm$ Gesture, $\pm$ Adjective) were filmed for each of the objects chosen in the norming study. Each participant saw 4 critical trials in each condition, along with 4 catch trials, for a total of 20 trials.

Similar to Experiment 1, counterbalancing was performed using 4 experimental lists, which participants were randomly sorted into by the Qualtrics survey platform. The questionnaire also presented trials within each list in a randomized order. The displays and the items within the displays were manipulated for counterbalancing, as in Experiment 1.



Figure 5: Sample forced choice task from Experiment 2 with the request, "Give me the short glass."

12

5.4 RESULTS. We analyzed the responses from the 79 participants who correctly answered at least 3 out of the 4 catch trials. Table 3 illustrates the aggregate behavior of the participants in each of the four conditions, showing the overall proportion of trials in which participants chose the adjective-contrast display. There was a clear trend: trials that included a lexical adjective resulted in adjective-contrast responses most of the time. Trials with no lexical adjective, on the other hand, also appeared to elicit active listener inference, but in favor of the noun-contrast display.

Next, we performed a logistic mixed effects regression using the glmer function in R. As in Experiment 1, item type and participant were taken into account as random effects.

As expected, there was a significant effect of lexical adjective on display choice ($p < 0.001$). Gestures, however, did not have a significant effect on response ($p = 0.2$). Surprisingly, this was true even for the -Adjective trials, in which the gestural content was not supported by a corresponding lexical modifier; there was no significant interaction effect between the gesture and adjective variables ($p = 0.8$). A closer examination of responses to individual gestures in the study does not reveal any gestures that lead to above-chance rates of the expected adjective-contrast responses in -Adjective trials . Interestingly, several gesture-noun pairs ([LONG] gloves, [BIG] lamp, [SHORT] string, [SMALL] teddy bear, and [BOW] tie) resulted in a roughly 50-50 split between the two display choices. In these trials, the gesture appears to negate the noun-contrast display preference motivated by the absence of the adjective. Though this pattern may be an indication of item-specific effects, the items in question all contain different modifier properties, and variation in responses across all items is not readily accounted for.

Analyzing contrastive inference responses at the level of the individual participant, we observe significant variation among individuals. We would expect most individuals to report contrastive inference by choosing the display with two similar objects more often in the +Adjective, –Gesture trials. However, some participants only reported the inference in about half of the trials, while others never reported contrastive inference. These outliers only made up 13% of participants; the remaining 77% reported contrastive inference in more than 75% of the trials in the condition.When we excluded the outliers and ran the regression only on data from the 77% of participants who responded to the task as predicted, there was still no significant interaction effect between gesture and linguistic cue ($p = 0.06$). There was, however, a weakly significant main effect of gesture ($p = 0.00764$).

5.5 DISCUSSION. Upon concluding that co-speech gesture does not participate in contrastive inferences like lexical modifierr, we are left to wonder, why would a speaker bother to expend valuable energy on gesture that is not intended to assert information? This result is still compatible with Gricean pragmatics if McNeill and Cooperrider are correct in that co-speech gesture is not intended to be communicative in the first place. However, disentangling that from not-at-issue

|  | Gesture | No gesture |
|---|---|---|
| Adjective | 84% | 82% |
| No adjective | 31% | 27% |

Table 3: Proportion of adjective-contrast responses by gesture and adjective condition

13

content more generally in this experimental setting is difficult. One way would be to investigate depictive gesture in sign languages; because both asserted and not-at-issue content in sign language is communicated gesturally, sign language data may help to answer what makes a gesture not-at-issue.

**6. General discussion and conclusions.** In the first experiment, two contrastive effects were pitted against each other: the pragmatic effect of modification, and the semantic effect of English prosodic focus. According to theories of focus (Rooth 1992; Roberts 2012), focusing a word expresses contrast between the spoken word and the unspoken alternatives. In the sentence "Give me the ADJ NOUN", moving prosodic focus from the adjective to the noun should then result in a completely different contrastive interpretation. Despite this prediction, our results suggested that the pragmatic effect of including an adjective overpowers the effect of prosodic focus, resulting in the same adjective-contrast display preference in both focus conditions. Focus did, however, significantly modulate the robustness of the resulting contrastive inference; adjective focus resulted in a greater preference for the adjective-contrast display.

Experiment 2 left the realm of verbal modification to investigate whether gestural modification also leads to contrastive inference. We found that co-speech gestures do not elicit or even modulate contrastive inferences; participant responses were entirely governed by the presence or absence of a verbal modifier. This result supports existing semantic analyses of co-speech gesture as not-at-issue; if a modifier is not part of the assertion, then a listener cannot take it to be crucial for interpretation. Thus, the line of pragmatic reasoning leading to contrastive inference is broken.

Overall, our offline measure of contrastive inference offers several practical advantages, including compatibility with video stimuli, and allowed us to study a range of questions. We successfully used it to replicate the results of previous online contrastive inference studies. In both experiments, simply adding an adjectival modifier before a noun had a significant effect on participant response with respect to inferred contrast. Our method's effectiveness at capturing contrastive inference leads us to an interesting conclusion: listeners are largely aware of and able to reason about their contrastive inferences. Though there is existing evidence from previous eye-tracking studies that contrastive inferences are generated incrementally, this offline task gets at the question of final interpretation, providing evidence for listener awareness of pragmatic reasoning.

**7. Conclusion.** In two experiments we investigated the expression of contrast, centered around two suprasegmental features: prosodic focus and co-speech gesture. We found in both cases that the presence or absence of an adjectival modifier results in robust contrastive inferences, and that this can be modulated by other factors such as focus. However, co-speech gestural modifiers do not support contrastive inferences in the same way as lexical modifiers. Given the simplicity of the experimental paradigm, we hope that our work inspires further investigations of this type of reasoning in other linguistic expressions.

**References**

Büring, Daniel. 2003. On d-trees, beans, and b-accents. *Linguistics and Philosophy* 26(5). 511–545. http://dx.doi.org/10.1023/A:1025887707652.

Cooperrider, Kensy. 2018. Foreground gesture, background gesture. *Gesture* 16(2). 176–202. http://dx.doi.org/10.1075/gest.16.2.02coo.

Eberhard, Kathleen M., Michael J. Spivey-Knowlton, Julie C. Sedivy & Michael K. Tanenhaus. 1995. Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research* 24(6). 409–436. http://dx.doi.org/10.1007/BF02143160.

Ebert, Cornelia & Christian Ebert. 2014. Gestures, demonstratives, and the attributive/referential distinction. Semantics and Philosophy in Europe. http://www.ilg.uni-stuttgart.de/mitarbeiter/ebert/papers/CECE-SPE2014.pdf.

Goldin-Meadow, Susan & Diane Brentari. 2015. Gesture, sign and language: the coming of age of sign language and gesture studies. *Behavioral and Brain Sciences* 40, e46. http://dx.doi.org/10.1017/S0140525X15001247.

Grice, H. P. 1975. Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and semantics 3: Speech arts*, 41–58. New York: Academic Press.

Grodner, Daniel & Julie C. Sedivy. 2011. The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (eds.), *The processing and acquisition of reference*, 239–271. Cambridge, MA: MIT Press. http://dx.doi.org/10.7551/mitpress/9780262015127.003.0010.

Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Levinson, Stephen C. 2000. *Presumptive meanings*. Cambridge, MA: MIT Press.

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

Pogue, Amanda, Chigusa Kurumada & Michael K. Tanenhaus. 2015. Talker-specific generalization of pragmatic inferences based on under- and over-informative prenominal adjective use. *Frontiers in Psychology* 6. 2035. http://dx.doi.org/10.3389/fpsyg.2015.02035.

Roberts, Craige. 2012. Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5. 69. http://dx.doi.org/10.3765/sp.5.6.

Rooth, Mats. 1992. A theory of focus interpretation. *Natural Language Semantics* 1(1). 75–116. http://dx.doi.org/10.1007/BF02342617.

Schlenker, Philippe. 2018. Gesture projection and cosuppositions. *Linguistics and Philosophy* 41(3). 295–365. http://dx.doi.org/10.1007/s10988-017-9225-8.

Sedivy, Julie C. 2003. Pragmatic versus form-based accounts of referential contrast: Evidence for effects of informativity expectations. *Journal of Psycholinguistic Research* 32(1). 3–23. http://dx.doi.org/10.1023/A:102192891.

Sedivy, Julie C., Michael K. Tanenhaus, Craig G. Chambers & Gregory N. Carlson. 1999. Achieving incremental semantic interpretation through contextual representation. *Cognition* 71(2). 109–147. http://dx.doi.org/10.1016/S0010-0277(99)00025-6.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard & Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268(5217). 1632–4. http://dx.doi.org/10.1126/science.7777863.