

ARTICLE

# Observers use gesture to disambiguate contrastive expressions of preference

Fey Parrill<sup>1\*</sup>  Jennifer Hinnell<sup>2</sup> , Grace Moran<sup>1</sup>, Hannah Boylan<sup>1</sup>, Ishita Gupta<sup>1</sup> and Aisha Zamir<sup>1</sup>

<sup>1</sup>Case Western Reserve University; <sup>2</sup> University of British Columbia

\*Corresponding author. Email: [fey.parrill@case.edu](mailto:fey.parrill@case.edu)

(Received 21 July 2020; Revised 17 August 2021; Accepted 19 August 2021)

## Abstract

We present two studies exploring how participants respond when a speaker contrasts two ideas, then expresses an ambiguous preference towards one of them. Study 1 showed that, when reading a speaker's preference as text, participants tended to choose whatever was said last as matching the speaker's preference, reflecting the recent-mention bias of anaphora resolution. In Study 2, we asked whether this pattern changed for audio versions of our stimuli. We found that it did not. We then asked whether observers used gesture to disambiguate the speaker's preference. Participants watched videos in which two statements were spoken. Co-speech gestures were produced during each statement, in two different locations. Next, an ambiguous preference for one option was spoken. In 'gesture disambiguating' trials, this statement was accompanied by a gesture in the same spatial location as the gesture accompanying the first statement. In 'gesture non-disambiguating' trials, no third gesture occurred. Participants chose the first statement as matching the speaker's preference more often for gesture disambiguating compared to non-disambiguating trials. Our findings add to the literature on resolution of ambiguous anaphoric reference involving concrete entities and discourse deixis, and we extend this literature to show that gestures indexing abstract ideas are also used during discourse comprehension.

**Keywords:** preference; contrast; gesture; embodiment; multimodality; reference tracking; ambiguity; discourse deixis; anaphora resolution

## 1. Introduction

Making choices and expressing preference is an inherent part of being human. Many choices, such as what to eat for breakfast or whether to have our morning coffee hot or iced, are ones that we make unconsciously and that have little long-term impact. But other choices have far-ranging effects, such as planning the direction of a business, or choosing a profession, where to live, or what to study. While the psychological processes involved in decision-making are largely the purview of the behavioral sciences (e.g.,

Brocas & Carrillo, 2014; Fellows, 2004; Kahneman & Tversky, 1981), how people communicate their preferences is fundamentally a linguistic phenomenon. Speakers convey their decision-making process and their final preferences through an array of word choices and morphosyntactic constructions appropriate to the context. For example, in response to the standard question when buying a morning coffee – “Is that for here or to go?” – the addressee might simply respond with “to go”. When selecting fruit at the grocery store, a speaker might state a preference by setting up a contrast using demonstrative pronouns *this* and *that*, as in “This one looks fresher than that one”. As speakers, we also move our bodies in ways that are aligned with speech. For example, when talking about the evaluation process for what university program to attend, a speaker might place both hands in a palm-up orientation as if holding objects, then perform a gesture that enacts the assessment of their relative weight.

Despite the staggering array of choices speakers make daily, the linguistic tools and coordinated kinesic expression of marking preference have received little explicit attention. Recent research in psychology, the cognitive sciences, cognitive linguistics, and gesture studies has shown that speech and the gestures that accompany speech, known as speech-accompanying or co-speech gesture (Kendon, 2004; McNeill, 1992, 2005; Zima & Bergs, 2017) are tightly coordinated. The two channels are very closely integrated both temporally and semantically (Church et al., 2014; Kelly et al., 2015; Kelly et al., 2010). We know that when people talk about weighing options and making decisions, they use gestures and other body movements (such as head tilts) to signal their mental construal of a contrastive set-up. For instance, the prototypical gesture for the idiomatic English expression *on the one hand ... on the other hand ...* is a bilateral gesture with each palm facing upwards (Hinnell, 2019). Comparing hypothetical options with *should I or shouldn't I ...*, speakers tend to enact similar bilateral gestures in combination with head tilts or simply perform a head tilt first to one side and then the other in coordination with the utterance (Hinnell, 2019). Research has also shown that speakers tend to produce gestures with alternate hands and anchor these in the gesture space around their body when they are comparing both concrete and abstract referents (Parrill & Stec, 2017). While these findings illuminate how speakers enact contrast, we know very little about the effect that these gestures have on the interlocutor, addressee, or observer. In the studies presented here, we address whether observers use a speaker's gesture to discern that speaker's preference between two possibilities.

A number of papers have shown that interlocutors track the spatial locations of gestures and use them to understand a discourse (Sekine & Kita, 2015, 2017; Sekine et al., 2015; So et al., 2009). These studies have focused on concrete entities (characters in a story, typically). Researchers have explained these findings by arguing that listeners create mental images of a discourse, encoding the spatial locations of entities in memory. That is, mental models of a discourse are embodied, in the sense that they include visual and motoric simulations (or partial reconstructions) of images and actions (Hostetter & Alibali, 2008, 2019). While claims about mental simulation have been most fully developed for the production side (see Hostetter and Alibali's Gestures as Simulated Action model), Sekine and Kita's active gestural discourse representation hypothesis makes a related argument for the comprehension side. According to these theories, observers should mentally simulate content they are encoding, thus linking the spatial location of a gesture to its referent in their mental model. Gestures in those locations can therefore help with retrieval. For example, if a speaker is describing a cat sitting on a fence, listeners will form a mental image of the locations of the cat and fence. Gestures can reinforce and elaborate that mental image.

But the same case has only recently (Hinnell & Parrill, 2020) been made for abstract entities (e.g., arguments and ideas), which do not have spatial locations. While theories of embodiment have argued that abstract content is also simulated, findings from experimental studies have been inconsistent (for a valuable summary of the general logic, see Diefenbach et al., 2013). If a speaker is talking about two relationships and gestures to each side, is that because she is actively maintaining a mental image with different spatial locations assigned to those relationships? Does her listener form a mental image that includes the locations of those abstract entities? Does that mental image become part of a discourse model in a way that facilitates comprehension? If observers use the spatial location of a gesture produced with an opinion to understand later discourse about that opinion, this finding would provide some evidence that we do indeed form such mental images and use gesture location to track abstract discourse referents. In short, the key contribution of our studies is the potential to extend arguments about embodied cognition beyond concrete reference into the domain of abstract reference.

In the next section, we contextualize the current study by reviewing relevant findings on anaphora resolution in sentence processing, on the use of gesture to track concrete referents in a discourse, and, finally, on the role of gesture in resolving ambiguous anaphora. We then describe two experiments that examine how observers respond to ambiguous expressions of preference and what role gesture plays in those responses. We close with implications for the field and areas of future research.

## 2. Background

In the current study, we explore how participants resolve ambiguous discourse reference in both a monomodal and multimodal context. The stimuli consist of scenarios with a pair of statements of preference (e.g., *Toni says pizza is best with pineapple. Marco really prefers more traditional pizza*), followed by one of the following ambiguous statements: *I agree with her/him/them, I'd prefer that, that makes more sense, that's more important to me, or s/he's/they're right*. In these statements, the personal or demonstrative pronouns are ambiguous in terms of which statement they refer to. Scenarios will be introduced in more detail below. We introduce them here to illustrate the relevance of several bodies of literature, namely, what we know about anaphora resolution in sentence processing, multi-modal (i.e., gestural) reference tracking, and the multimodal depiction of contrast.

### 2.1. Sentence processing and anaphora resolution

Monomodal studies have shown that the resolution of anaphoric reference (e.g., pronouns that refer back to a previous noun phrase) relies on a very wide set of linguistic variables in the speech context. These include the linguistic salience or accessibility of the referent, which have been shown to be related to subject position (Gernsbacher & Hargreaves, 1988; Nappa & Arnold, 2014), givenness (Gundel et al., 1993), and thematic roles (e.g., the stimulus role; Arnold, 2001).

Subject, or first-mention bias, is exemplified in the sentence pair *My mom went shopping with my sister. She bought shoes*, in which the ambiguous pronoun *she* will most frequently be interpreted as referring back to *my mom*. There is also evidence for a recent mention bias over clause boundaries. For example, Arnold (2010)

demonstrated in a corpus study that “reference to information mentioned in recent clauses is more likely to be pronominal than less recent information” (Arnold 2010, p. 189). In a later study, Arnold summarized these findings by saying that listeners build assumptions based on their linguistic experience. Arnold argues “that speakers tend to continue talking about recently mentioned entities especially subjects” (Arnold et al., 2018, p. 42).

While these findings are from studies using concrete referents (e.g., characters), there is also an extant body of work on the resolution of abstract antecedents in anaphora resolution. This is often referred to as discourse deixis, defined as anaphoric reference that involves reference to abstract entities such as events and states (Asher, 1993). Discourse deixis is exemplified in the clause pair *Max destroyed his leaf collection last night. That was dumb* (Hegarty et al., 2001, p. 138), where *that* refers to the whole event in the first clause. Much of this research relates specifically to discourse status (e.g., given vs. new information; Gundel et al., 1993), rather than to differences between abstract vs. concrete entities, which is a distinction we make here.

Both the recent mention bias found for concrete entities and the research on discourse deixis are important factors in the studies presented here. We argue that the recent-mention bias explains the overall patterns in both Study 1 and Study 2. We specifically argue that gesture is being used to override this bias for some trials. Discourse deixis is relevant in that the demonstrative pronoun *that* appears in one of the (ambiguous) constructions in our stimuli: *I guess I agree more with that.*

## 2.2. Multimodal reference tracking in production

Studies of multimodal (i.e., embodied) behavior are also relevant to anaphora resolution since speakers and signers make use of gesture space (the area directly around their torso where co-speech gesture generally occurs) to anchor referents in space. For example, when participants retold stories they were shown in a series of vignettes, speakers used the location of their gestures to identify characters in the story, in addition to specifying them in their speech (So et al., 2009). Sekine and Kita (2017) found that if a speaker produced gestures in particular locations for story characters, observers used those locations to resolve reference, even when no further gestures were produced (see also Sekine & Kita, 2015; Sekine et al., 2015). They termed this the ‘active gestural discourse representation hypothesis’, which proposes that gesture serves as part of a mental model of discourse entities and events in short-term memory. Parrill and Stec (2017) found that gesture space is also used in a consistent way for gestures about abstract referents.

The use of space around the body to anchor ideas is also a known feature of sign languages. In outlining the spatial mapping strategies of American Sign Language, Winston (1995, 1996) describes how signers construct a spatial map by anchoring entities on either side of the signing space through pointing in sequence to each side. The signer then continues to refer to the two entities throughout the discourse by pointing to the two spaces in which these entities are anchored (Janzen, 2012). Some research has explicitly investigated differences between the use of referential space in sign versus spoken language. For example, in an investigation of the marking of referential space in both German Sign Language and in co-speech gesture accompanying German in elicited narratives, Perniss and Ozyürek (2015, p. 36) found that “the configuration of referent locations in sign space and gesture space corresponds

in an iconic and consistent way to the locations of referents in the narrated event<sup>9</sup>. They go on to describe how the use of space is modified in specific contexts for sign and co-speech gesture modalities depending on whether referents were re-introduced in a discourse or maintained across consecutive clauses.

### 2.3. Multimodal reference tracking in perception

The research cited above demonstrates how speakers consistently produce gestures and use gestures referentially. However, it is also important to know what information the interlocutor is gaining from these gestures. We know that addressees pay attention to co-speech gestures and use semantic information to improve understanding (e.g., Beattie & Shovelton, 2002). When observing gestures that do not match someone's speech, gestural details are incorporated into our representations (Broaders & Goldin-Meadow, 2010; McNeill et al., 1994). In more recent research, a study of pointing gestures with non-present referents (i.e., pointing to empty space, rather than at an object) found that inconsistent use of gesture space impaired language processing (Gunter et al., 2015). In an experimental set-up featuring a mismatch paradigm, participants watched a video in which a woman being interviewed anchored referents such as Donald Duck and Mickey Mouse in gesture space. In scenarios in which the pointing gesture accompanied a target word that was not consistent with a previously established location in gesture space, participants exhibited a larger N400, a neural measure of language processing that has a higher amplitude when an element is unexpected or harder to process. Finally, a qualitative study of dyadic interaction showed that the speaker and addressee co-construct spatial reference even across longer periods of time (Stec & Huiskes, 2014).

Specific to perception studies of reference resolution, when addressees track a referent in discourse, they appear to be sensitive to the presence of gesture; however, studies of anaphora resolution diverge widely in the effect found. In a study based on a series of reaction time experiments, Debreslioska and colleagues (2019) explored localizing anaphoric gestures in gesture-congruent and gesture-incongruent conditions. In their reaction time tasks, participants watched videos of narratives involving two (or more) characters. Participants were instructed that the first-mentioned entity was the main protagonist and were shown stimuli with video such as the following narrative: *There was a woman. And her husband couldn't repair the engine of his car by himself. So, the woman decided to call/write to her brother. He should come to help him out*, in which gestures to left/right gesture space aligned with the underlined segments. Debreslioska and colleagues predicted that participants would be faster to respond to the question *Did the main protagonist call for help?* in the gesture congruent condition – a prediction that was not, in fact, borne out in their two studies. While this study tracked only one referent through the short narrative, others (Nappa & Arnold, 2014; Sekine & Kita, 2017) have examined contrasted entities, i.e., by localizing two referents in opposite gesture spaces, as our stimuli do. Nappa and Arnold (2014) reported no difference between the incongruent and no-gesture conditions in their study, while Sekine and Kita (2017) showed slower reaction times in the incongruent condition in their study.<sup>1</sup>

<sup>1</sup>See Debreslioska et al. (2019) for an in-depth discussion of diverging results in perception studies of localizing gestures.

In an experiment using taped narratives in which the speaker used a pronoun that could refer to either of the previously mentioned same-gender referents, and in which a gesture co-occurred with one of the two possible referents, Goodrich Smith and Hudson Kam (2012) replicated the first-mention bias discussed above. However, in a finding replicated in Study 2 presented here, when a localizing gesture co-occurred with the second-mentioned referent, participants chose that referent more often than a referent with no gesture. (However, they still chose the first referent more often overall.) An important distinction is that Goodrich Smith and Hudson Kam (2012) and the other perception studies cited here involve concrete entities, i.e., characters in a narrative, rather than gestures that localize abstract referents (e.g., opinions, statements) as in the studies presented here.

#### 2.4. Gestures depict mental structures of contrast

Given our focus on contrastive statements of preference, studies on the multimodal expression of contrast are relevant here as well. A recent corpus study of expressions of contrast in North American English using the Red Hen multimedia archive (Steen & Turner, 2013; [www.redhenlab.org](http://www.redhenlab.org)) showed that speakers consistently use a subset of hand shapes anchored on either side of the body to gesture contrast (Hinnell, 2019, 2020). This was true of comparisons of objects and abstract notions in the real world, as well as in comparisons of two hypothetical situations (e.g., *If X / if Y*, and *should I / shouldn't I*). Findings suggest that contrast is systematically expressed through lateral and symmetrical manual gestures (e.g., bilateral palm-up open-hand). Enactments of contrast that were aligned with hypothetical expressions (i.e., in the irrealis domain, marked linguistically by conditionals and modal verbs) featured either head tilts (i.e., first to one side then the other) or head tilts in coordination with manual gesture (Hinnell, 2019, 2020). The following examples (drawn from Hinnell, 2019) illustrate these patterns. The prototypical expression of contrast using the palm-up open-hand gesture form (PUOH; Müller, 2008) in conjunction with *on the one hand / on the other hand*, is shown in Figure 1. In Figure 2, two abstract entities (*Republican news network* and *Republican presidential candidates*) are anchored on alternate sides of the speaker's gesture space using a container gesture, in which the hands are seen to hold a metaphorical object between them, while Figure 3 shows a head tilt accomplishing the same general function of placing objects on either side of the gesture space.

While these examples show contrast being expressed through phrasal expressions and antonymic noun and adverbial phrases, contrast can also be expressed at a discourse level through concessive markers such as *anyhow*, *anyways*, *but*, *though*, and *nevertheless*, and phrasal markers with the same function such as *at the same time* (Biber et al., 1999; cf. Barth, 2000; Couper-Kuhlen & Thompson, 2000; Ford, 2000). The corpus study referenced earlier (Hinnell, 2019) showed that this type of discourse-level contrast is also marked in the body, via body lean, gesture, and head tilts. Given their role in marking contrast, concessive markers were used for the stimuli in the studies presented here.

Why do speakers gesture, lean, or tilt their heads to different sides, and produce gestures as though they are holding or placing objects to one side or the other? According to conceptual metaphor theory (Johnson, 1987, 2017; Lakoff & Johnson, 1980, 1999), humans conceptualize decision-making or preference as balance or



S: Apparently they like him on both sides. On the one hand he believes in climate change so they consider him a Liberal.

G: right hand PUOH



On the other hand he doesn't believe in abortion so they consider him a Republican...

left hand PUOH

Fig. 1. Palm-up open-hand gesture: 'weighing options'.



S: Republican news network versus

G: bimanual container gesture to left of body



frontrunner Republican presidential candidate

bimanual container gesture to right of body

Fig. 2. Container gesture: 'placing options in space'.



S: Should I be flattered?

G: gaze upwards, head tilt (R)



Should I be outraged?

gaze upwards head tilt

Fig 3. Head tilt: 'placing options in space'.

weight to one side or the other. Some experimental support for this claim comes from a study showing that manipulating participants' body position impacts their preferences (Dijkstra et al., 2012). Gestures reflect imagery and motor representations (Hostetter & Alibali, 2008, 2019), thus we may see lateral gestures during expressions of contrast because humans conceptualize the ideas that they are talking about as objects that can be placed or held (see for example Parrill, 2008). On the other hand, PUOH gestures are so common in discourse that some authors have argued that they are conventionalized (Müller, 2008), rather than reflecting active metaphoric imagery or mental simulation.

The literature cited here shows how speakers and signers organize sign and gesture space to anchor concrete references. It also suggests that observers perceive gestures as conveying meaningful information, though the findings from perception studies of anaphora resolution are mixed. Our study goes beyond the question of whether speakers use gesture to help resolve reference for concrete referents (characters, objects, spatial locations). We ask whether observers will use gesture to resolve reference when the entity being talked about is abstract, i.e., What is the effect of gesture on an observer's perception of a speaker's preferences for one possibility over another? The referents in our study are discourse deictic elements (Levinson, 1983; Webber, 1991). That is, rather than a person or object or spatial location, the referents that observers must track are what the speaker has said. There is good reason to expect that gesture will be valuable here as well, based on previous reference-tracking studies and corpus studies of naturalistic behavior. However, to our knowledge the only study showing that observers use gesture to resolve ambiguous pronouns that have abstract antecedents is an experiment conducted by the authors (Hinnell & Parrill, 2020), which we comment on in the discussion (Section 5).<sup>2</sup> For concrete referents, the argument is that listeners form a mental image of what they are hearing, thus incorporating spatial locations into that mental image (Perniss & Özyürek, 2015), as the active gestural discourse representation hypothesis (Sekine & Kita, 2017) would presumably also argue. That listeners do something similar for abstract referents needs to be established with experimental data. As noted in the introduction (Section 1), while people gesture in ways that maintain the spatial relationships between concrete entities, presumably because they form mental images of those relationships (Hostetter & Alibali, 2008, 2019), only limited evidence is available for abstract entities (Hinnell & Parrill, 2020). Language users do seem to use space consistently for abstract reference (Parrill & Stec, 2017), but these patterns and those found in work by Hinnell may be the result of convention rather than mental imagery or simulation. If observers use the spatial location of a gesture produced with an opinion to understand later discourse about that opinion, this finding would support the claim that listeners are also forming mental images that include the locations of abstract entities. Furthermore, even the findings from perception studies regarding concrete referents are unclear.

We present two studies that explore how participants respond to ambiguity when a speaker is expressing a preference. The first study establishes a baseline by asking how speakers respond in a monomodal environment (i.e., reading a speaker's words as text). Based on pilot work, we anticipated that, in the absence of any information,

<sup>2</sup>The Hinnell and Parrill (2020) study was carried out after the work reported here, but was published first. Hinnell and Parrill (2020) is a partial replication and extension of these studies.

participants will choose the last thing read or heard as the speaker's preference (a recent mention bias). In the second study, we investigate the role of gesture in disambiguating the speaker's preference.

### 3. Study 1: How do observers respond to ambiguous statements of preference?

Study 1 asked how participants respond to an ambiguous expression of preference when reading text.

#### 3.1. Method

##### 3.1.1. Participants

Thirty students from a private university in the northeastern US took part in the study for course extra credit. Twenty identified as female, none identified as non-binary. The mean age was 20, and the range was 18–21. All were native speakers of American English.

##### 3.1.2. Materials

We created 59 scenarios that contained the following elements:

1. An attitude about a topic (this will be referred to as the A statement)
2. A concessive
3. A differing attitude about the topic (referred to as the B statement)
4. A hedge indicating uncertainty
5. An ambiguous statement that used a pronoun to indicating a preference for either the A or B (referred to as the preference statement).

For example, one scenario was as follows: *Toni says pizza is best with pineapple* (A statement). *On the other hand* (concessive), *Marco really prefers more traditional pizza* (B statement). *I'm not sure* (hedge), *but I think he's right* (preference statement). The preference statement is ambiguous because *he* could refer to either Toni or Marco. We used five different concessive expressions: *but*, *but also*, *but then*, *but at the same time*, and *on the other hand*. These concessives were selected because a corpus study of contrast (Hinnell, 2020) indicated that they are frequently used in contrastive speech. The preference statements contained one of the following: *I agree with her/him/them*, *I'd prefer that*, *that makes more sense*, *that's more important to me*, or *s/he's/they're right*.

##### 3.1.3. Procedure

Participants were presented with these scenarios in random order using Qualtrics survey software. After an informed consent screen, participants were presented with a scenario as text. After reading the scenario, participants responded to a question asking for their judgment about the stimulus speaker's preference on a new screen. The exact question was matched to the preference statement, so that, for example, a preference statement ending with *I'd prefer that* would be followed by the question *What does the speaker prefer?* In the example scenario above, the question would be

*Who does the speaker think is right?* Participants chose between options matched to the scenario, such as *pineapple pizza* and *traditional pizza*. Options were presented horizontally, and their locations were random (thus, the option appearing to the left was random for each trial so that the choice options did not necessarily match the spatial location of the A and B statements from the participant's point of view). The A and B statements were not counterbalanced. That is, for the previous example, all participants saw pineapple pizza in the A statement, and traditional pizza in the B statement.

Second, a new screen asked participants to respond to the question *What is your personal opinion/preference?*, and they were presented with the same options as in the previous question (e.g., *pineapple pizza*, *traditional pizza*). The location of options was also randomized with respect to location. Participants answered these questions for each of the 36 scenarios, then completed a demographic survey asking about age and sex/gender (male, female, non-binary).

The key dependent variable for our study is the participant's judgment about the stimulus speaker's preference. We refer to this as 'stimulus speaker's preference'. The variable describing what participants chose for their own preference is referred to as 'participant preference'. The A and B statements were not counterbalanced, so we collected this information in case it turned out that participants tended to prefer one of the two options above chance (e.g., traditional pizza is more popular).

Because our pilot work indicated that participants would choose the B statement more often than the A statement for stimulus speaker's preference, we coded the choice of the A statement as 0 and the choice of the B statement as 1. We hypothesized that we would see a proportion higher than .50 for B statement responses.

### 3.2. Results

One scenario was removed due to a data collection error leaving a total of 58. Data have been uploaded to Open Science framework and can be found here at <<https://osf.io/fzhwj/>>. Data were analyzed using R version 4.0.0 (R Core Team, 2020). Utility packages used for data manipulation, cleaning, and analysis include tidyverse (Wickham et al., 2019; Wickham & Henry, 2020), dplyr (Wickham et al., 2020), broom (Robinson et al., 2020), lme4 (Bates et al., 2015), and afex (Singmann et al., 2020).

Figure 4 shows a boxplot for the proportion of participants choosing the B statement across the two dependent variables. Our analyses are designed to answer two separate questions: (1) Do participants choose the B statement at a rate higher than chance; and (2) Do participants choose the B statement more often for the stimulus speaker's preference compared to their own preference?

#### 3.2.1. Do participants choose the B statement at a rate higher than chance?

We carried out a Shapiro–Wilk normality test, finding that stimulus speaker data were non-normal ( $W = 0.90624$ ,  $p = .01197$ ), so a non-parametric test should be used. Participant preference data were roughly normally distributed ( $W = 0.96242$ ,  $p = .3566$ ), so a parametric or a non-parametric test could be used. We opted to use a Wilcoxon test to compare the proportion choosing B to chance (.5). We calculated the proportion of B choices for each participant, created a column that has .50 (the chance level) for each participant, and compared the two columns. We did this for

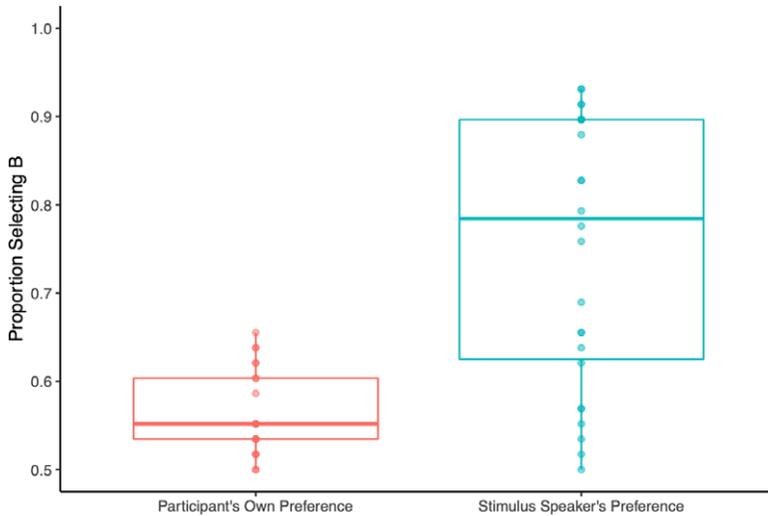


Fig. 4. Boxplot showing proportion choosing the B statement.

both dependent variables separately. The Wilcoxon signed rank test for stimulus speaker preference versus chance (0.5) showed a statistically significant difference ( $V = 414$ ,  $p < .00001$ ). Individuals did not appear to be randomly selecting 'A' vs. 'B' when asked for the stimulus speaker's preferences, with participants choosing B an average of 71% of the time ( $M = 0.71$ ,  $SD = 0.18$ ). The Wilcoxon signed rank test for the participants' own preference versus chance (0.5) also showed a marginally significant difference ( $V = 324$ ,  $p = .06$ ). Individuals did not appear to be randomly selecting 'A' vs. 'B' by chance when asked for their own preferences, with participants choosing B an average of 54% of the time ( $M = 0.54$ ,  $SD = 0.06$ ).

Consistent with findings from the recent mention bias literature, participants chose the B statement more often for stimulus speaker preference (about 70% of the time). For participant's own preference, because the A and B statement were not counterbalanced, we saw a proportion much closer to chance (54%), indicating that some participants preferred A and some preferred B. This difference was also statistically different from chance, indicating a bias towards the B statement overall. This pattern is considered in greater detail in the general discussion (Section 5).

### 3.2.2. Do participants choose the B statement more often for stimulus speaker's preference compared to their own preference?

In our study design, 30 participants responded to 58 scenarios. For each scenario, the participant answered two questions: What does the speaker prefer? (stimulus speaker preference) and What is your personal opinion/preference? (participant's preference). To see whether the rate of choosing the B statement was any different for the two questions, we ran a mixed-model logistic regression (logit) model with participant and scenario as random effects, and fixed effect of 'Whose Preference' (stimulus speaker or participant's own). The dependent variable was choosing the B statement as preference, with a score of 0 if the individual chose the A statement, and 1 if they chose the B statement. The model was fit using maximum likelihood (Laplace

Approximation), with a random effect of scenario (Intercept Variance = 0.595,  $SD = 0.77$ ) and a random effect of participant (Intercept Variance = 0.1941,  $SD = 0.44$ ). The fixed effect of ‘Whose Preference’ was statistically significant ( $\beta = 0.87$ ,  $SE = 0.077$ ,  $z = 11.263$ ,  $p < .0001$ ), with participants being 1.4 times more likely to respond ‘B’ when asked the stimulus preference than when asked their own preference. The model has a residual error of 4174.2 on 3476 degrees of freedom; an Akaike’s information criterion (AIC) of 4182.2; and fixed intercept values as follows:  $\beta = 0.16970$ ,  $SE = 0.13950$ ,  $z = 1.216$ ,  $p = .224$ .

Finally, although our research question was not concerned with the influence of different concessives, we include brief analyses exploring the impact of concessive in the Appendix. Broadly speaking, the pattern was the same regardless of which concessive appeared in the scenario.

### 3.3. Discussion

Taken together, these analyses indicated that participants tended to think the stimulus speaker preferred the B option (whatever was expressed last) in the absence of any disambiguating information. These findings are consistent with the literature on recent mention as a factor in reference resolution (e.g., Arnold, 2010). They also had a slight bias towards the B statement for their own preference, which we will comment on in greater detail in the general discussion (Section 5).

## 4. Study 2: How does gesture affect observers’ responses to ambiguous statements of preference?

Having established how participants respond to ambiguous text scenarios, in Study 2 we asked whether participants will use gesture to disambiguate these kinds of scenarios. We first created an audio version of our scenarios. Comparing audio to video will allow us to rule out the possibility that intonation explains some part of our findings. In our main study, we presented participants with video versions of scenarios in which the speaker performed gestures with the A and B statements. In some trials, a gesture also occurred with the preference statement. This gesture was always in the same location as the gesture co-occurring with the A statement. We asked: Given that participants tend to choose the B statement for the stimulus speaker’s preference, are they *less likely to do so* if the ambiguous preference statement is accompanied by a gesture in a location previously established as the ‘A statement location’? (That is, if the gesture is performed in the same location as a previous gesture that co-occurred with the A statement.)

### 4.1. Method

#### 4.1.1. Participants

150 participants were recruited using the online data collection platform Amazon Mechanical Turk. Mechanical Turk data have been shown to be comparable to data collected in academic research studies, but the Mechanical Turk population is more diverse in age, education, and race/ethnicity than most typical university research populations (Buhrmester et al., 2011). Participants were required to be within the

United States to take part and were compensated with three dollars (the study took about half an hour).

#### 4.1.2. Materials

We identified the 36 scenarios from Study 1 that had the highest proportion of participants choosing the B statement for the stimulus speaker's preference. These ranged from .68 to .98. Scenarios (and proportions choosing B for both variables) can be found in the Appendix.

We first recorded audio for each scenario. A female research assistant was trained on the intonational properties of the scenarios by the second author. Peak prosodic emphasis was on the subject noun phrase of the statements (e.g., *Toni, Marco*), and the ambiguous noun phrase of the preference statement (e.g., *he's*). After this training, the research assistant was instructed to read each scenario as naturally as possible. We next recorded video for each scenario. The same research assistant was instructed to sit in a comfortable posture, and to perform several scenarios as naturally as possible. Scenarios were performed in two different ways, a gesture-disambiguating version (GD) and a gesture non-disambiguating version (GND).

Both versions used palm-up open-hand gestures. These were performed with the A and B statements. The research assistant performed versions with the left hand first and with the right hand first. For the GND version, the speaker sat still during the preference statement. For the GD version, the speaker performed a final palm-up open-hand gesture with the preference statement. The final gesture always occurred in the location where the A statement gesture had been performed. For example, if the first gesture was on the left, the final gesture would be performed on the left as well.

We selected the best videos to create four categories: (1) right hand first, left hand second, final gesture with right hand; (2) left hand first, right hand second, final gesture with left hand; (3) right hand first, left hand second, no third gesture; and (4) left hand first, right hand second, no third gesture. These were used to create final videos for all scenarios. Using Final Cut Pro, we matched audio files to these four different videos to create two stimulus lists. Stimulus lists counterbalanced trial type (GD, GND) across specific scenarios. That is, a scenario about pizza preference would be matched with GD video for list 1 and GND video for list 2. We also counterbalanced which hand gestured first for each scenario across lists. This use of lists was intended to minimize the chances that specific properties of the scenarios would interfere with our results. Due to an error (discussed below) scenarios were assigned unevenly, with 17 assigned as GD trials, and 19 assigned as GND trials. We next used Final Cut Pro to align gesture strokes (the effortful, meaningful portion of a gesture; McNeill, 1992) with the subject noun phrases (e.g., *Toni, Marco*), by adding or removing slices of audio or video. We edited the video so that the gesture was held until the end of the statement, then edited the video so that the gesture retraction (return to rest) occurred after the end of the statement. We also edited the video so that the stroke of the final gesture (for GD stimuli) was aligned with the ambiguous noun phrase (e.g., *he's*), held for the rest of the statement, and the retraction occurred after the end of the statement.

We used Final Cut Pro to blur the speaker's face and upper shoulders so that mouth movements did not reveal the fact that audio and video had been edited, as



Fig. 5. Example stimulus.

shown in Figure 5. We also did this masking so that variability in facial expressions and head movements would not affect participants' judgments. There was some variation in intonational contours and in the research assistant's posture across each scenario. This was desirable, as it made the scenarios appear more natural. Once the videos were completed, we exported the audio tracks. These audio tracks were used for an audio-only condition.

To recapitulate: the outcome of the audio/video editing was to create two versions of each scenario, with scenarios paired to GD and GND videos in a counterbalanced way. Within those categories, scenarios were paired in a counterbalanced way with videos in which the right versus left hand was used first. Audio and video were carefully aligned to preserve the systematicity of auditory and gestural cues, and the audio-only condition used exactly the same audio as the gesture condition. Participants saw only one version of each scenario (either as a GD or a GND video).

#### 4.1.3. Procedure

Participants received a link to access the study and were randomly directed to one of three versions: audio only, gesture version list 1, or gesture version list 2. The surveys were presented using Qualtrics survey software. After an informed consent screen, participants were presented with a stimulus (audio or video), then answered the questions described above (the participant's judgment about the stimulus speaker's preference and the participant's own preference). Each element (audio/video, first question, second question) was presented on a separate screen. As with Study 1, the location (left, right) of the options was randomized. This ensured that, for example, when a gesture was performed on the speaker's left in a video, the location of the option associated with the co-occurring NP would be on the left for approximately half the participants and on the right for approximately half. Participants answered these questions for each of the 36 scenarios, then completed a demographic survey asking about age, sex (male, female, non-binary), race/ethnicity, and what the participant thought the study was about.

## 4.2. Results

Data have been uploaded to Open Science framework and can be found here at <<https://osf.io/fzhwj/>>. Data were analyzed using R version 4.0.0. Details on specific packages can be found in the results section for Study 1 (Section 3.2). Data were removed if the participant did not answer all questions or completed the study too quickly to have done the task correctly (7 participants were removed; thus the final number of participants was as follows: audio only = 45, gesture version list 1 = 46, and gesture version list 2 = 52). Demographic details are presented in the Appendix. The dependent variables for study 2 are the same as for study 1, namely stimulus speaker's preference and participant preference.

### 4.2.1. Audio results

Figure 6 shows a boxplot for the proportion of participants choosing the B statement across the two dependent variables. The audio version of the scenarios was intended to rule out the possibility that intonation explained some part of potential differences we saw when gesture was added. Because the audio condition is not comparable to the video condition (it has no disambiguating and non-disambiguating trials), we analyzed it separately.

As with study 1, our analyses answer two different questions: (1) Do participants choose the B statement at a rate higher than chance? and (2) Do participants choose the B statement more often for the stimulus speaker's preference compared to their own preference? To answer the first question, we carried out a Shapiro–Wilk normality test, finding that stimulus speaker audio data were non-normal ( $W = 0.927, p = .0078$ ), so a non-parametric test should be used. Participant preference audio data were roughly normally distributed ( $W = 0.982, p = .732$ ), so a parametric or a non-parametric test could be used. As with Study 1, we used a Wilcoxon test to compare the proportion choosing B to chance (.50). We calculated the proportion of B choices for each participant, created a column that had .50 (the chance level) for

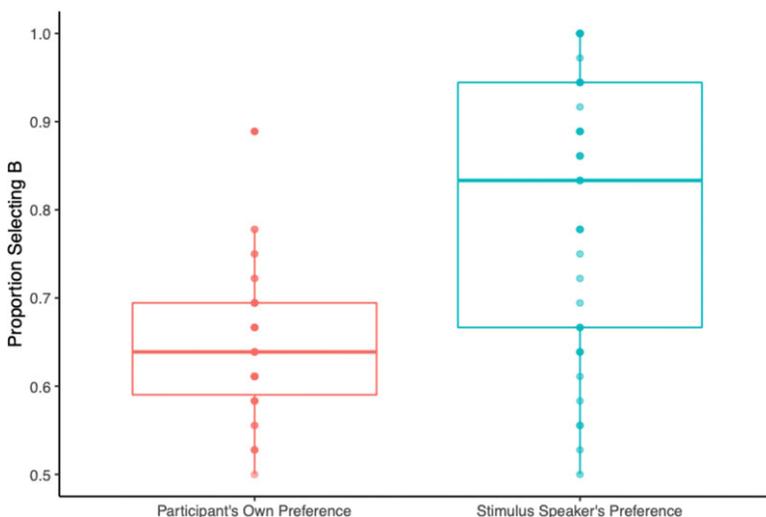


Fig. 6. Proportion choosing the B statement for audio condition.

each participant, and compared the two columns. We did this for both dependent variables separately. The Wilcoxon signed rank test for stimulus speaker preference vs. chance (0.5) showed a statistically significant difference ( $V = 972$ ,  $p < .00001$ ). Individuals did not appear to be randomly selecting 'A' vs. 'B' by chance when asked the stimulus speaker's preferences, with participants choosing B an average of 77% of the time ( $M = 0.77$ ,  $SD = 0.18$ ). The Wilcoxon signed rank test for the participants' own preference vs. chance (0.5) also showed a statistically significant difference ( $V = 923$ ,  $p < .0001$ ). Individuals did not appear to be randomly selecting 'A' vs. 'B' by chance when asked for their own preferences, with participants choosing B an average of 54% of the time ( $M = 0.54$ ,  $SD = 0.10$ ).

To answer our second question (Is the rate of choosing the B statement different for the two questions?), we ran a mixed-model logistic regression (logit) model with participant and scenario as random effects, and fixed effect of 'Whose Preference' (stimulus speaker or participant's own). We included participant and scenario as random intercepts with the goal of increasing the generalizability of the findings outside the specific group of people and scenarios (see, e.g., Bresnan, 2007; Quené & van den Burgh, 2008). We did not fit random slopes because there is a limited number of possible slopes for these data, thus no theoretically motivated reason for including this term. The dependent variable was choosing the B statement as the preference, with a score of 0 if the individual chose the A statement, and 1 if they chose the B statement. The model was fit using maximum likelihood (Laplace Approximation), with a random effect of scenario (Intercept Variance = 0.32,  $SD = 0.568$ ) and a random effect of participant (Intercept Variance = 0.434,  $SD = 0.659$ ). The fixed effect of 'Whose Preference' was statistically significant ( $\beta = 0.869$ ,  $SE = 0.083$ ,  $z = 10.354$ ,  $p < .0001$ ), with participants being 1.38 times more likely to respond 'B' when asked for the stimulus preference compared to when asked for their own preference. The model has a residual error of 3633.3 on 3236 degrees of freedom; an AIC of 3641.3.

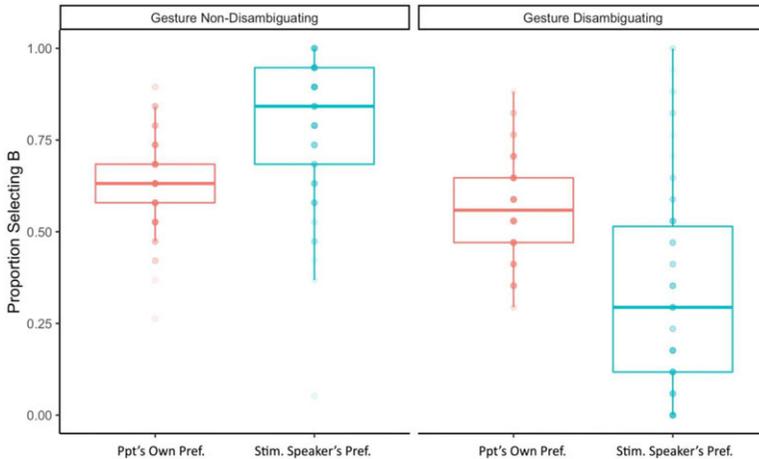
#### 4.2.2. Video results

Because the only purpose of our two stimulus lists was to counterbalance trial type across different scenarios, we collapsed over the two lists. Thus, 98 participants responded to 36 scenarios. As with previous analyses, we started by asking whether participants chose the B statement at a rate higher than chance for the two questions (what does the speaker prefer, stimulus speaker preference, and what is your personal preference, participant preference). We examined this separately for the two kinds of trial types (GD, GND). Again, we carried out a Shapiro–Wilk normality test for each grouping, which generally indicated that data were non-normal, suggesting a non-parametric test should be used. Test statistics for each individual test can be found in the Appendix. We then carried out Wilcoxon signed rank tests, all of which showed a statistically significant difference from chance. Individuals did not appear to be randomly selecting 'A' vs. 'B'. Table 1 shows means, SDs,  $V$  values, and  $p$  values for individual tests, and Figure 7 shows a boxplot by response and trial type.

Our main research question was whether the rate of choosing the B statement for stimulus speaker preference was lower for gesture disambiguating trials compared to gesture non-disambiguating trials. To see whether the rate of choosing the B statement was different for the two questions, we ran a mixed model logistic regression (logit) model with participant and scenario as random effects (again

**Table 1.** Results of Wilcoxon tests comparing each grouping to chance

	Mean	SD	V	P
Stimulus speaker, GND	.79	.17	4713.5	< .0001
Stimulus speaker, GD	.31	.27	850	< .0001
Participant preference, GND	.63	.12	4554.5	< .0001
Participant preference, GD	.56	.14	3602.5	< .0001



**Fig. 7.** Boxplot for video data.

fitting no random slopes). Fixed effects were trial type (GD vs. GND) and ‘Whose Preference’ (stimulus speaker or participant’s own), and their interaction. The dependent variable was choosing the B statement as a preference, with a score of 0 if the individual chose the A statement, and 1 if they chose the B statement. The model was fit using maximum likelihood (Laplace Approximation). Table 2 shows the model summary for change scores, including *z* values and *p* values. With GD trials set as the comparison, we found a statistically significant difference in the selection of the B statement across trial type. With stimulus speaker preference set as the comparison, we found a statistically significant effect of ‘Whose Preference’, and a significant interaction between trial type and ‘Whose Preference’, AIC = 8390.5 on 7050 degrees of freedom.

Table 3 shows a comparison between the model with the interaction, one with no interaction, models with only one fixed effect (trial type or ‘Whose Preference’) and an empty model. The decrease in AIC across the models suggests that the full model performs best. Finally, we used the emmeans package to perform pairwise comparisons for Trial Type and ‘Whose Preference’. Table 4 shows these contrasts. Because participants were more likely to choose the B statement for all trial types, there are multiple significant contrasts. The most theoretically relevant contrasts are those between the different trial types (GD, GND) for stimulus speaker preference, which is statistically different, versus that same difference across trial types for participant’s own preference, which is not statistically different.

**Table 2.** Logistic regression model for summary for trial type and whose preference, comparison group in parentheses

	Estimate	Std Error	Z value	Pr(> z )
(Intercept)	-0.7897	0.11	-7.136	< .0001
Trial Type GND	2.1829	0.0944	23.106	< .0001
Whose Preference (Participant Preference)	1.10719	0.07553	14.658	< .0001
Trial Type (GND) × Whose Preference (Participant Preference)	-1.94454	0.10863	-17.901	< .0001

**Table 3.** Model comparison, Study 2 video data

	Full Model	No Interaction	Trial Type	Whose Preference	Empty Model
Gesture Disambiguating	-0.238*** (0.09)	-1.155*** (0.07)	-1.153*** (0.07)		
Whose Preference	0.837*** (0.08)	-0.165*** (0.05)		-0.158*** (0.05)	
Interaction	-1.945*** (0.11)				
(Intercept)	0.556*** (0.11)	1.019*** (0.10)	0.935*** (0.10)	0.463*** (0.12)	0.383*** (0.12)
Observations	7056	7056	7056	7056	7056
Log Likelihood	-4189	-4356	-4361	-4493	-4497
Akaike Inf. Crit.	8391	8722	8730	8993	9001
Bayesian Inf. Crit.	8432	8756	8757	9021	9021

Note: \*p, \*\*p, \*\*\*p < .01.

**Table 4.** Pairwise comparisons

Comparison		Estimate	SE	Z ratio	P value
GD Ppt's own pref vs. GND Ppt's own pref		-0.2559	0.0995	-2.573	0.0605
GD Ppt's own pref vs. GD Stim spkr pref		-0.0108	0.0734	-0.147	1
GD Ppt's own pref vs. GND Stim spkr pref		-0.2667	0.0995	-2.682	0.0439
GND Ppt's own pref vs. GD Stim spkr pref		0.2965	0.0996	2.978	0.0174
GND Ppt's own pref vs. GND Stim spkr pref		0.0406	0.0711	0.571	1
GD Stim spkr pref vs. GND Stim spkr pref		-0.3073	0.0995	-3.087	0.0121

### 4.3. Discussion

The results of Study 2 suggest that participants chose the B statement at a rate higher than chance for both stimulus speaker preference, and the participant's own preference, when listening to audio scenarios. This pattern remained true when presented with video scenarios in which the speaker performed gestures only while voicing the A and B statements. The pattern was different when the speaker performed a third gesture while voicing the preference statement. Because this gesture occurred in the same spatial location as the previously uttered A statement, in this case only participants were more likely to choose the A statement for stimulus speaker's preference (chose A about 70% of the time on average). This pattern suggests that observers maintained an association between the side of space where a gesture had previously occurred during the A statement, and the option voiced along with that

gesture. Observers then used that association to interpret the ambiguous preference statement.

## 5. General discussion

This project is grounded in a corpus study showing that expressions of contrast are consistently accompanied by gestures anchored on either side of the body (Hinnell, 2019). We extended these findings by showing that the occurrence of a gesture changes the way an observer interprets the speaker's preference. That is, an ambiguous expression of preference without gesture is interpreted differently than one with gesture. This is consistent with a body of research indicating that observers track the spatial locations of gestures and integrate them into their representations of what the speaker is saying. We were also able to show that spatial location tracking occurs with abstract references (discourse-deictic elements), as other studies have shown for concrete referents. The results of Study 1 indicate that, when responding to an ambiguous scenario presented as text, participants will choose the last thing they encountered as the speaker's preference, about 70% of the time. Our participants also tended to choose the last thing they read as their own preference, about 54% of the time. When presented with audio scenarios (Study 2), participants also chose the B statement for the stimulus speaker's preference more often than chance. The rate for audio was slightly higher (77% of responses) compared to text. Participants also continued to choose the B statement at a rate slightly higher than chance for their own preference (54%).

When presented with video scenarios, we saw a similar pattern if there was no gesture occurring with the preference statement. For those trials (gesture non-disambiguating, GND), participants chose the B statement about 80% of the time on average for the stimulus speaker's preference. They chose B for their own preference about 60% of the time. However, when a gesture occurred in the same spatial location as a previously uttered referent, participants chose the B statement only 30% of the time as the stimulus speaker's preference. When gesture occurred with the preference statement in a location previously associated with the A statement, participants chose the B statement for their own preference 63% of the time on average.

Taken together, these findings are consistent with the literature on recent mention as a factor in reference resolution (e.g., Arnold, 2010). People will assume that the last thing said is what the speaker is referring to. But, crucially, gesture can be used to weaken that assumption.

### 5.1. Limitations

Several limitations arose from design choices we made in creating experimentally controlled videos. Our participants were presented with a series of statements from the same (face-less) speaker, and their response is doubtless rather different from the way gesture is integrated into mental representations during dynamic, real-life interactions. Although only 20 to 30 percent of our participants mentioned gesture or body language when asked what they thought the study was about, the artificial nature of the stimuli presumably directed attention to gesture in ways that only generally reflect typical processing. We did not perform any analyses asking whether

guessing the purpose of the study predicted response, but this could be explored in future work. Next steps might vary the handshape of the gestures, the prosody, involve more speakers and naturalistic settings, and generally increase the extent to which the stimuli represent real discourse. In addition, because we chose to match audio and video features (gesture strokes and peak prosodic emphasis for noun phrases), the audio for gesture disambiguating and non-disambiguating versions of each scenario differed slightly. These differences were millisecond differences in the length of pauses, rather than differences in intonation. It is possible that these minute audio differences actually explain our findings.<sup>3</sup> We think this is unlikely: the variability in audio editing was not constant across trials because the length of each noun phrase varied. That is, as far as we know there is no systematic feature in the audio that could be tied to a theoretically relevant phenomenon. That said, we cannot rule it out.

An unexpected finding was that participants also chose the B statement as their own preference more frequently than chance. This occurred for all versions of our scenarios, ranging from 54% to 63%. A simple explanation for this phenomenon is that, because we did not counterbalance the order of the statements, the more popular choice (on average) was B. For example, traditional pizza was always the B statement in our scenarios, and perhaps more of our participants prefer traditional pizza. Based on responses from Study 1, the average for participant preference across all scenarios was 60% (see proportions by scenario in the Appendix). This may mean that the overall tendency to choose B was amplified. This is not problematic for our findings, because the same logic would not apply to the selection of stimulus speaker preference. In Study 2, we saw that both fixed effects (trial type, 'Whose Preference') predicted the choice of A or B, and their interaction also predicted the choice of A or B. But while the contrast between the two trial types was statistically different for stimulus speaker preference, it was not for participants' own preference.

Although we feel that this is a good explanation for the patterns we observed, it is also possible that a recent mention bias impacted participants' choices for their own preference, possibly because this was an experimental situation with zero stakes. That is, when pressing a button to decide which kind of pizza you prefer, absent a strong feeling, you may choose whatever you heard last. Our study was not designed to separate these possibilities. Future work could ensure the study begins with a set of normed options with preferences balanced, or the statements could be counterbalanced. But the neutrality of the scenarios does raise an interesting question about whether the relative plausibility of the two options shapes the role played by gesture. That is, if listeners find it highly implausible that the speaker could actually take the 'A' position, will gesture still perform a disambiguating function when it occurs in the 'A statement' location? We recently extended our paradigm by comparing neutral to moral scenarios (for example, a neutral scenario such as buying pink sandals versus an evaluation of human cloning, which the majority of Americans consider to be immoral: Gallup, 2017). We found that participants did not respond differently to morally charged scenarios (Hinnell & Parrill, 2020). When a speaker's gesture co-occurred with a statement about a very unpopular position (e.g., human cloning is good), a gesture in the same spatial location co-occurring with an ambiguous statement was still used to disambiguate the statement.

<sup>3</sup>We thank an anonymous reviewer for pointing out this issue.

## 5.2. Future directions

The key finding of our study is that the presence of a gesture changed how participants interpreted a speaker's statements. A number of questions remain. First, our study does not explain *why* participants choose the last thing they encountered as the stimulus speaker's preference. Our argument is that this is the general function of concession (that is, concession allows speakers to negate the first discourse element and align themselves with the second), and that this pattern reflects a general tendency to choose the most recent referent in resolving an ambiguous pronoun (e.g., Arnold, 2010).

A second question raised by our findings concerns the fact that, even when seeing a disambiguating gesture, participants did not always choose the statement that the gesture occurred with. Participants still chose the B statement around 30 percent of the time (decreasing from around 70% without disambiguating gesture). This pattern indicates that the information coming from the gestural modality is balanced against the discourse function of a concessive (the speaker believes whatever she says last), and recent mention biases. Previous studies have also found that gesture interacts with other factors in a more nuanced way than might be expected. Debreslioska and colleagues (2019) discuss the fact that results of studies on localizing gestures for concrete referents do not always converge. This is perhaps not surprising given that gesture is expected to be the less-attended channel of information, thus it is unlikely to completely override other factors such as the recent mention bias.

In addition, it is important to understand the role of gesture in statements of preference that are not ambiguous. Several studies have found no facilitatory effect of gesture on response time in gesture-congruent conditions (Sekine & Kita, 2017), but Gunter and Weinbrenner (2017) suggest that gestures must be valid cues for a task in order to facilitate comprehension. One possible extension of the study presented here is to examine whether gestures facilitate comprehension of statements of preference that are not ambiguous and to investigate the degree to which observers attend to gestures in non-ambiguous linguistic contexts (i.e., those that are disambiguated entirely in the speech to prevent an unnatural focus on gesture).

Gestures that mark preference display an embodied grounding in our experience of contrast in the real world. We weigh ideas or objects against each other and we refer to these in the ongoing discourse by indexing these options in the gesture space, most frequently to either side of our body (Hinnell, 2019, Parrill & Stec, 2017). While gesture has been referred to as an 'unwitting' window into the mind of the speaker/gesturer (McNeill, 1992), in this study, we reversed the question. Rather than exploring how preferences are enacted by speakers, we investigated the effect of a speaker's gesture on an observer. Findings showed that gestures can shift an observer's interpretation of speaker preference. These findings go beyond previous studies of cohesive and referential gestures by using abstract referents (i.e., discourse deictic elements) rather than concrete ones. More broadly, these findings provide supporting evidence for the tight integration of speech and gesture (Church et al., 2014; Kelly et al., 2015; Kelly et al., 2010), for the activation of gesture throughout a discourse, and for the role of gesture in an observer's comprehension of discourse (Sekine & Kita, 2017). The findings also provide perception evidence that humans conceptualize decision making or preference as balance or weight to one side or the other (e.g., Hinnell, 2019, 2020; Johnson, 1987, 2017; Turner, 1991),

thereby extending arguments about embodied cognition beyond concrete reference into the domain of abstract reference. While some proponents of embodied theories of cognition suggest there is no great difference between a concrete object (a cat) and an abstract object (the right way to stay mentally sharp), the literature on the comprehension of abstract language has not always found the same patterns for the two types of language (see Diefenbach et al., 2013). In addition, the arguments about visual and motor simulation (Hostetter & Alibali, 2008, 2019) – and how such putative processes connect to co-speech gestures – are not as straightforward for abstract speech. Abstract referents do not have spatial locations until a conceptualizer creates them. It is important and worthwhile to show that gestures accompanying speech about ideas can serve similar discourse functions to gestures accompanying speech about objects.

**Acknowledgements.** We thank Raedy Ping for advice on statistical procedures and Nicole Baumgartner for assistance with stimulus preparation.

**Conflicts of Interest.** We have no conflicts of interest to declare.

## References

- Arnold, J. E. (2001). The effect of thematic roles on pronoun use and frequency of reference continuation. *Discourse Processes* 31(2), 137–162.
- Arnold, J. E. (2010). How speakers refer: the role of accessibility. *Language and Linguistic Compass* 4(4), 187–203.
- Arnold, J. E., Strangmann, I. M., Hwang, H., Zerkle, S. & Nappa, R. (2018). Linguistic experience affects pronoun interpretation. *Journal of Memory and Language* 102, 41–54.
- Asher, N. (1993). *Reference to abstract objects in discourse*. Amsterdam: Springer.
- Barth, D. (2000). ‘That’s true, although not really, but still’: expressing concession in spoken English. In E. Couper-Kuhlen & B. Kortmann (eds), *Cause – condition – concession – contrast: cognitive and discourse perspectives* (pp. 411–438). Berlin: De Gruyter.
- Bates, D., Maechler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67(1), 1–48.
- Beattie, G. & Shovelton, H. (2002). An experimental investigation of some properties of individual iconic gestures that mediate their communicative power. *British Journal of Psychology* 93(2), 179–192.
- Biber, D., Johansson, S., Leech, G., Conrad, S. & Finegan, E. (eds) (1999). *Longman grammar of spoken and written English*. London: Pearson Education Limited.
- Bresnan, J. (2007). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. In S. Featherston & W. Sternefeld (eds), *Roots: linguistics in search of its evidential base* (pp. 77–96). New York: De Gruyter Mouton.
- Broaders, S. C. & Goldin-Meadow, S. (2010). Truth is at hand: how gesture adds information during investigative interviews. *Psychological Science* 21(5), 623–628.
- Brocas, I. & Carrillo, J. D. (2014). Dual-process theories of decision-making: a selective survey. *Journal of Economic Psychology* 41, 45–54.
- Buhrmester, M., Kwang, T. & Gosling, S. D. (2011). Amazon’s Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science* 6(1), 3–5.
- Church, R. B., Kelly, S. & Holcombe, D. (2014). Temporal synchrony between speech, action and gesture during language production. *Language and Cognitive Processes* 29(3), 345–354.
- Couper-Kuhlen, E. & Thompson, S. (2000). Concessive patterns in conversation. In E. Couper-Kuhlen & B. Kortmann (eds), *Cause – condition – concession – contrast: cognitive and discourse perspectives* (pp. 381–410). Berlin: De Gruyter.
- Debreslioska, S., van de Weijer, J. & Gullberg, M. (2019). Addressees are sensitive to the presence of gesture when tracking a single referent in discourse. *Frontiers in Psychology* 10(1775), 1–14.

- Diefenbach, C., Rieger, M., Massen, C. & Prinz, W. (2013). Action-sentence compatibility: the role of action effects and timing. *Frontiers in Psychology* 4(272), 1–13.
- Dijkstra, K., Eerland, A., Zijlman, J. & Post, L. S. (2012). How body balance influences political party evaluations: a Wii balance board study. *Frontiers in Psychology* 3(536), 1–8.
- Fellows, L. K. (2004). The cognitive neuroscience of human decision making: a review and conceptual framework. *Behavioral and Cognitive Neuroscience Reviews* 3(3), 159–172.
- Ford, C. (2000). The treatment of contrasts in interaction. In E. Couper-Kuhlen & B. Kortmann (eds), *Cause – condition – concession – contrast: cognitive and discourse perspectives* (pp. 283–311). Berlin: De Gruyter.
- Gallup (2017). *Values & beliefs poll*. Retrieved from <<https://news.gallup.com/poll/210542/americans-hold-record-liberal-views-moral-issues.aspx>>.
- Gernsbacher, M. A. & Hargreaves, D. (1988). Accessing sentence participants: the advantage of first mention. *Journal of Memory & Language* 27(6), 99–717.
- Goodrich Smith, W. & Hudson Kam, C. L. (2012). Knowing ‘who she is’ based on ‘where she is’: the effect of co-speech gesture on pronoun comprehension. *Language and Cognition* 4(2), 75–98.
- Gundel, J. K., Hedberg, N. & Zacharski, R. (1993). Cognitive status and the form of referring expressions in discourse. *Language* 69, 274–307.
- Gunter, T. C. & Weinbrenner, J. E. D. (2017). When to take a gesture seriously: on how we use and prioritize communicative cues. *Journal of Cognitive Neuroscience* 29(8), 1355–1367.
- Gunter, T. C., Weinbrenner, J. E. D. & Holle, H. (2015). Inconsistent use of gesture space during abstract pointing impairs language comprehension. *Frontiers in Psychology* 6(80), 1–10.
- Hegarty, M., Gundel, J. & Borthen, K. (2001). Information structure and the accessibility of causally introduced referents. *Theoretical Linguistics* 27(2/3), 163–186.
- Hinnell, J. (2019). The verbal-kinesic enactment of contrast in North American English. *American Journal of Semiotics* 35(1/2), 55–92.
- Hinnell, J. (2020). *Language in the body: multimodality in grammar and discourse*. Unpublished doctoral dissertation, University of Alberta.
- Hinnell, J. & Parrill, F. (2020). Gesture influences resolution of ambiguous statements of neutral and moral preferences. *Frontiers in Psychology* 11(587129).
- Hostetter, A. B. & Alibali, M. W. (2008). Visible embodiment: gesture as simulated action. *Psychonomic Bulletin & Review* 15(3), 495–514.
- Hostetter, A. B. & Alibali, M. W. (2019). Gesture as simulated action: revisiting the framework. *Psychonomic Bulletin & Review* 26, 721–752.
- Janzen, T. (2012). Two ways of conceptualizing space: motivating the use of static and rotated vantage point space in ASL discourse. In B. Dancygier & E. Sweetser (eds), *Viewpoint in language* (pp. 156–176). New York: Cambridge University Press.
- Johnson, M. (1987). *The body in the mind: the bodily basis of meaning, imagination, and reason*. Chicago, IL: University of Chicago Press.
- Johnson, M. (2017). *Embodied mind, meaning, and reason: how our bodies give rise to understanding*. Chicago, IL: University of Chicago Press.
- Kelly, S. D., Healey, M., Özyürek, A. & Holler, J. (2015). The processing of speech, gesture, and action during language comprehension. *Psychonomic Bulletin & Review* 22(2), 517–523.
- Kelly, S. D., Özyürek, A. & Maris, E. (2010). Two sides of the same coin: speech and gesture mutually interact to enhance comprehension. *Psychological Science* 21(2), 260–267.
- Kendon, A. (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. Chicago, IL: University of Chicago Press.
- Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh: the embodied mind and its challenge to western thought*. New York: Basic Books.
- Levinson, S. C. (1983). *Pragmatics*. Cambridge: Cambridge University Press.
- McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. Chicago, IL: University of Chicago Press.
- McNeill, D., Cassell, J. & McCullough, K. E. (1994). Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction* 27(3), 223–237.
- McNeill, D. (2005). *Gesture and thought*. Chicago: University of Chicago Press.
- Müller, C. (2008). What gestures reveal about the nature of metaphor. In A. Cienki & C. Müller (eds), *Metaphor and gesture* (pp. 219–245). Amsterdam: John Benjamins.

- Nappa, R. & Arnold, J. E. (2014). The road to understanding is paved with the speaker's intentions: cues to the speaker's attention and intentions affect pronoun comprehension. *Cognitive Psychology* 70, 58–81.
- Parrill, F. (2008). Form, meaning, and convention: a comparison of a metaphoric gesture with an emblem. In A. Cienki & C. Müller (eds), *Metaphor and gesture* (pp. 195–217). Amsterdam: John Benjamins.
- Parrill, F. & Stec, K. (2017). Gestures of the abstract: Do speakers use space consistently and contrastively when gesturing about abstract concepts? *Pragmatics and Cognition* 24(1), 33–61.
- Perniss, P. & Ozyürek, A. (2015). Visible cohesion: a comparison of reference tracking in sign, speech, and co-speech gesture. *Topics in Cognitive Science* 7(1), 36–60.
- Quené, H. & van den Burgh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language* 59(4), 413–425.
- R Core Team. (2020). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. Online <https://www.R-project.org/>.
- Robinson, D., Hayes, A. & Couch, S. (2020). *Broom: convert statistical objects into tidy tibbles*. R package version 0.7.3. Online <https://CRAN.R-project.org/package=broom>.
- Sekine, K. & Kita, S. (2015). Development of multimodal discourse comprehension: cohesive use of space in gesture. *Language, Cognition, and Neuroscience* 30(10), 1245–1258.
- Sekine, K. & Kita, S. (2017). The listener automatically uses spatial story representations from the speaker's cohesive gestures when processing subsequent sentences without gestures. *Acta Psychologica* 179, 89–95.
- Sekine, K., Sowden, H. & Kita, S. (2015). The development of the ability to semantically integrate information in speech and iconic gesture in comprehension. *Cognitive Science* 39(8), 1855–1880.
- Singmann, H., Bolker, B., Westfall, J., Aust, F. & Ben-Shachar, M. (2020). *afex: Analysis of Factorial Experiments*. R package version 0.28-0. Online <https://CRAN.R-project.org/package=afex>.
- So, W. C., Kita, S. & Goldin-Meadow, S. (2009). Using the hands to identify who does what to whom: gesture and speech go hand-in-hand. *Cognitive Science* 33(1), 115–125.
- Stec, K. & Huiskes, M. (2014). Co-constructing referential space in multimodal narratives. *Cognitive Semiotics* 7(1), 31–59.
- Steen, F. F. & Turner, M. (2013). Multimodal construction grammar. In M. Borkent, B. Dancygier & J. Hinnell (eds), *Language and the creative mind* (pp. 1–20). Stanford, CA: CSLI Publications.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211 (453–211).
- Turner, M. 1991. *Reading minds*. Princeton, NJ: Princeton University Press.
- Webber, B. L. (1991). Structure and ostension in the interpretation of discourse deixis. *Language and Cognitive Processes* 6(2), 107–135.
- Wickham, H. et al. (2019). Welcome to the tidyverse. *Journal of Open Source Software* 4(43), 1686.
- Wickham, H., François, R., Henry, L. & Müller, K. (2020). *dplyr: a grammar of data manipulation*. R package version 0.8.5. Online <https://CRAN.R-project.org/package=dplyr>.
- Wickham, H. & Henry, L. (2020). *tidyr: tidy messy data*. R package version 1.0.2. Online <https://CRAN.R-project.org/package=tidyr>.
- Winston, E. A. (1995). Spatial mapping in comparative discourse frames. In K. Emmorey & J. S. Reilly (eds), *Language, gesture and space* (pp. 87–114). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Winston, E. A. (1996). *Spatial mapping in ASL discourse*. Paper presented at the CIT 11th National Convention, Little Rock, Arkansas.
- Zima, E. & Bergs, A. (2017). Towards a multimodal construction grammar. *Linguistics Vanguard* [Special Issue], 3(s1).

## Appendix

### Study 2 scenarios

Table A1 presents the scenarios used in Study 2. Columns three and four show the mean proportion choosing B. These were selected from the Study 1 scenarios because they had the highest proportion of participants choosing the B statement for the stimulus speaker's preference.

Table A1. Study 2 scenarios

#	Scenario	Stim Sprk Pref	Partic Pref
1	Some people are gonna vote based on party. On the other hand, some people are gonna vote based on specific issues. I care about both things, but I guess that's more important to me.	0.96	0.84
2	My mom likes manicures because everyone can see her nails, even in the winter. On the other hand, she also says pedicures are more relaxing. I'm not sure I want to do either, but I guess I'd prefer that.	0.76	0.12
3	My brother just does a load of laundry whenever he needs to. But at the same time My sister does all her laundry once a week. Laundry is not my favorite chore, but I guess I'd prefer that.	0.84	0.36
4	My brother never checks the weather, he doesn't really care. On the other hand my dad checks every day so he's never surprised. I don't worry about it a lot, but I think I agree with him.	0.88	0.96
5	I know a lot of people who don't cook cause it's too much trouble. On the other hand, plenty of people are like it's not that hard to make spaghetti. I can see it both ways, but I think that makes more sense.	0.92	0.72
6	My grandma always puts off calling the doctor as long as she can. On the other hand, my grandpa calls right away just in case. They both have good reasons, but that makes more sense to me.	0.92	0.6
7	Some people say if you read on a screen you don't read as carefully. On the other hand some people say it doesn't make a difference. It probably depends on a lot of things, but that makes more sense to me.	0.72	0.36
8	Some people think if a kid doesn't do well in a class it's because they're not working hard. On the other hand, some people think teachers don't have enough training to teach certain things. It's tough, but I guess I agree with them.	0.88	0.68
9	Lauren likes those automatic hand dryers because they are more energy efficient. On the other hand, Kira hates them because they are so loud. I get both their points, but I think she's right.	0.72	0.4
10	Jason likes to take the train to games so we don't have to park On the other hand Sammy likes to drive because it's a million times faster I don't have strong feelings, but his argument makes more sense to me.	0.68	0.52
11	Mary says you have to toast the bread for a peanut butter and jelly sandwich. On the other hand, Cara says a real PB&J can never be toasted. Both points are valid, but I guess I agree more with that.	0.76	0.72
12	Some people think if you give poor people money without conditions they will spend it on drugs or alcohol. On the other hand, many people think they'll spend it on food and rent and education. I don't know, but I guess I agree more with that.	0.84	0.64
13	My little brother's not on facebook because he thinks it's a waste of time. On the other hand, my other brother says he can't do job networking without it. They both make sense, but I guess I agree with him.	0.8	0.56
14	Sarah says you shouldn't call yourself a Christian if you don't ever go to church. But at the same time, Maria says lots of Christians have a more personal relationship to God. Hard to say, but I guess that makes more sense.	0.72	0.6
15	Molly thinks people shouldn't say they have food allergies just because they don't want to eat certain things. But at the same time, she also	0.72	0.4

(Continued)

Table A1. (Continued)

#	Scenario	Stim Sprk Pref	Partic Pref
	knows people say they are allergic when they aren't technically, because they really don't feel good if they eat dairy or whatever. I don't know, but I guess I agree more with that.		
16	Some doctors are saying vaping is safe because there's no risk of cancer like with tobacco. But at the same time, others are saying vaping is not safe because we don't have any research about the chemicals. I don't plan on taking it up, but I guess that makes more sense.	0.84	0.96
17	Josh would rather drive for an hour to go for a hike because he wants to be away from the city. But Fernando says just go to a local park and spend your time outside, not in a car. Not sure, but his argument makes more sense to me.	0.76	0.36
18	A lot of people are very specific in what kind of person they will date. But other people are very open to different kinds of people. I'm not sure, but I guess I'd prefer that.	0.8	0.64
19	Some people cross at the crosswalk because they think it's safer. But some people cross when nobody's coming because they think that's safe. Could go either way, but I guess that makes more sense.	0.72	0.72
20	Mark always goes to pick up the pizza because it's way faster. But miles would rather wait and not have to leave the house. They both make good points, but I guess I agree with him.	0.76	0.6
21	My brother gets mad at my dad for eating too much chocolate. But my dad says that chocolate is better for you than beer. Mmmm. I'm not sure, but his argument makes more sense to me.	0.88	0.76
22	Julie was looking to buy pink sandals because they're more exciting. But at the same time Christa said get black ones cause they go with everything. Good points, but I think she's right.	0.8	0.8
23	Some people say driving a fast car is the best feeling in the world. But a lot of people think a regular car is just as good since there's a speed limit. For me it's a mix, but I think they're right.	0.72	0.68
24	Experts say that workers should get overtime pay to protect workers' rights. But at the same time they say it's not fair because managers don't get overtime. Both points are valid, but I guess I agree more with that.	0.76	0.16
25	A lot of people are having a hard time making ends meet. On the other hand plenty of people say that our economy is doing just fine. I don't know, but I think they're right.	0.72	0.28
26	A lot of candidates memorize portions of their speech instead of speaking naturally. On the other hand some candidates think speaking freely sounds better. I don't know how much it matters, but I think they're right.	0.8	0.48
27	My grandmother always said listening to music keeps you young. On the other hand, my grandfather says that peace and quiet does. Either way, both are important, but I think that makes more sense.	0.76	0.24
28	With climate change some people are planning for the possibility of extinction. On the other hand, a lot of people say there is still time to change our trajectory I don't know, but I think they're right.	0.8	0.8
29	My dad says that doing crosswords helps him stay mentally sharp. But my brother says reading a lot is the best way to improve mental skills. I can see it both ways, but I think that makes more sense.	0.8	0.92

(Continued)

Table A1. (Continued)

#	Scenario	Stim Sprk Pref	Partic Pref
30	Lawn care experts say you should leave the clippings on the lawn because it's good for the grass. On the other hand, people say you should bag the clippings because it looks messy. I can see it both ways, but I guess I agree with them.	0.76	0.56
31	Mohammed thinks living in a condo building is safer because you're not on the ground floor. On the other hand, Chris says that in a house you can get out faster if there's a fire. I'm not sure, but I guess I agree with him.	0.72	0.68
32	Most farmers think of their animals as livestock. But non-farmers think of animals as pets. I'm not sure, but I guess I agree more with that.	1	0.8
33	José thinks shopping at independent stores is better for the environment. On the other hand, Ahmil thinks big box stores are better because you make less trash from packaging. I can see it both ways, but I think he's right.	0.76	0.4
34	Alexandra thinks riding the bus across the country is the best way to see it. On the other hand, she also says she'd have more time to do stuff if she flew instead. I can see it both ways, but I guess I'd prefer that.	0.88	0.92
35	A lot of people think learning a second language is the best way to experience another culture. On the other hand, some people say traveling to a country is the best way to experience the culture. I can see it both ways, but I think that makes more sense.	0.8	0.8
36	Toni says pizza is best with pineapple. On the other hand Marco really prefers more traditional pizza. Mmmm. I'm not sure, but I think he's right.	0.84	0.8

### Analyses by concessive

We did not design our scenarios to test the impact of different concessives but, to explore their possible impact we used a mixed model logistic regression (logit) model with participant as random effect, and fixed effects of 'Whose Preference' (stimulus speaker or participant's own) and concessive. The dependent variable was choosing the B statement as preference, with a score of 0 if the individual chose the A statement, and 1 if they chose the B statement. With 'but' as the comparison level, we found a marginally significant effect for 'but also', but no other differences as a function of concessive. For Study 2, we chose the Study 1 scenarios that had the highest proportion of participants choosing the B statement for the stimulus speaker's preference. None of these scenarios used 'but also'.

Table A2. Model summary for choice of A versus B by concessive

Fixed effects	Estimate	Std Error	z value	Pr(> z )
(Intercept)	0.127	0.116	1.091	0.2752
Whose preference (Stimulus speaker pref)	0.778	0.072	10.690	<.0001
But also	-0.272	0.148	-1.838	0.0661
But then	-0.026	0.162	-0.162	0.8709
but at the same time	0.094	0.1268	0.741	0.4586
on the other hand	0.05567	0.10	0.554	0.5793

### Shapiro–Wilk test results for Study 2 video version

To assess the normality of the video data, we carried out Shapiro–Wilk tests. They are included here to make the main text more readable.

**Table A3.** Study 2 Shapiro–Wilk test results

	<i>W</i>	<i>p</i>
Stimulus speaker, GND	.899	< .0001
Stimulus speaker, GD	.916	< .0001
Participant preference, GND	.975	.0646
Participant preference, GD	.971	.031

### Study 2 demographic information

Demographic data were collected to allow for the possibility of exploratory analyses that could be used in designing future studies. We had no theoretical questions about age, race, or sex/gender so do not conduct any analyses here.

**Table A4.** Study 2 demographic information

	Audio	List 1	List 2
Female*	42%	43%	38%
Age, average and range	38.2, 24–68	36.3, 22–57	36.3, 22–67
Race**			
White	77.78%	80.44%	76.92%
Asian	8.89%	8.70%	5.77%
Black or African American	11.11%	17.39%	17.31%
American Indian or Alaskan Native	0%	0%	1.92%
Native Hawaiian or Other Pacific Islander	0%	2.17%	0%
Two or More	0%	6.52%	1.92%
Other	2.22%	2.17%	1.92%
Spanish/Hispanic/Latino	8.89%	13.04%	17.31%
Fluent in English only	88.89%	97.83%	92.31%
Mentioned Gesture / Body Language / Action	0%	30.44%	23.08%

Notes: \* Choices were Male/Female/Other, no participant identified as other; \*\*Participants were asked “Do you identify as ...?”.

**Cite this article:** Parrill, F., Hinnell, J., Moran, G., Boylan, H., Gupta, I. & Zamir, A. (2022). Observers use gesture to disambiguate contrastive expressions of preference *Language and Cognition* 14: 19–46, <https://doi.org/10.1017/langcog.2021.20>