Language and Cognition (2026), 18, e43, 1–27
doi:10.1017/langcog.2026.10086

ARTICLE

Adapting pragmatic interpretation in multiparty conversations: mutual exclusivity inferences across speaker and context

Edmundo Kronmüller[1] and Ernesto Guerra[2,3]

[1]Escuela de Psicología, Pontificia Universidad Católica de Chile, Santiago, Chile; [2]Centro de Investigación Avanzada en Educación, Instituto de Educación, Universidad de Chile, Santiago, Chile and [3]Departamento de Lingüística, Facultad de Filosofía y Humanidades, Universidad de Chile, Santiago, Chile
Corresponding author: Edmundo Kronmüller; Email: ekr@uc.cl

(Received 11 July 2025; Revised 02 April 2026; Accepted 23 April 2026)

Abstract

Listeners flexibly adapt utterance interpretation based on speaker characteristics, but less is known about how this works in multiparty conversations, particularly for pragmatic inferences like Mutual Exclusivity (MEI) – the tendency to map novel terms onto unnamed referents. MEI may reflect a flexible, speaker-sensitive process or a rigid default. This study tested whether listeners modulate MEI based on speaker-specific referential consistency in multiparty conversations. Two eye-tracking experiments were conducted (Exp 1: N = 32, English; Exp 2: N = 44, Spanish), where participants followed instructions from two speakers. In experimental conditions, one speaker was consistent (reused labels), while the other was inconsistent (sometimes relabeled objects). Control conditions featured two consistent speakers. Results showed robust and early MEI across conditions. No speaker-specific adaptation occurred, but Experiment 1 showed context-general adaptation: exposure to an inconsistent speaker within a conversation reduced MEI overall. Findings suggest that in high-demand communicative contexts, MEI operates as a strong default heuristic, largely resistant to speaker-level inconsistency. Evidence for sensitivity to broader contextual reliability was observed in Experiment 1 but did not replicate in Experiment 2, leaving open whether MEI exhibits limited contextual flexibility or whether the initial finding reflected design-specific factors.

Keywords: eye-tracking; multiparty conversation; mutual exclusivity; pragmatic inferences; speaker specificity

1. Introduction

One key characteristic of utterance interpretation in conversational contexts is its flexibility. In going from conventional meanings to context-sensitive interpretations, listeners can rely on different pieces of information that determine what they

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

understand. One of the main sources of information is speakers and their characteristics. People are heterogeneous in many dimensions, such as age, language proficiency, shared knowledge, personality and tone of voice (among many others), and listeners make interpretations considering these differences. However, everyday conversations commonly involve more than two interlocutors. Are listeners able to track the source of variability – specific speakers – and adjust their interpretations accordingly? The goal of the present article is to determine whether listeners adapt their interpretations to specific speakers based on their characteristics, specifically, their *referential consistency* in multiparty conversations. We take as a case study Mutual Exclusivity Inferences (MEI) upon hearing a brand-new definite description in the context of established precedents in referential communication.

Mutual Exclusivity Inference is a powerful cognitive strategy that helps resolve referential ambiguity by assuming that each word corresponds to a unique referent (Kachergis et al., 2012; Markman & Wachtel, 1988). When encountering a novel word, listeners typically exclude familiar objects with known labels and map the novel term onto previously unnamed objects. Mutual Exclusivity has been extensively studied in the context of language acquisition, where it enables children to constrain possible word meanings and efficiently expand their vocabulary (Lewis et al., 2020). However, MEI is not limited to development; it also functions in adult communication as a pragmatic principle that guides referential interpretation. In conversational contexts, when a speaker introduces a new expression after having established a referential precedent for an object, listeners tend to avoid mapping the new expression to the same object. This results in interference when no suitable alternative referent is available (Kronmüller & Barr, 2007; Metzing & Brennan, 2003; Wu et al., 2024).

The question of whether listeners can flexibly adapt their pragmatic inferences based on speaker-specific characteristics is particularly pertinent to MEI. Previous research has demonstrated that listeners modulate other types of pragmatic inferences, such as scalar and contrastive inferences, in response to speaker characteristics. For example, Grodner and Sedivy (2011) found that listeners suspend their contrastive interpretation of scalar adjectives when interacting with speakers described as having 'language and social problems'. Similarly, Ryskin et al. (2019) showed that listeners adjust their interpretative processes based on evidence of a speaker's reliability in using prenominal adjectives contrastively. These findings suggest that listeners maintain sophisticated models of speaker-specific linguistic behavior and adjust their interpretations accordingly. However, it remains unclear whether and how speaker-specific characteristics modulate the application of MEI, particularly in multiparty conversations where listeners must track multiple speakers simultaneously.

The literature on partner-specific adaptation in language processing provides mixed evidence regarding the time course and mechanisms of such adaptations. Some research suggests that speaker-specific information is processed in two distinct stages: an early, automatic retrieval of speaker-related episodic memories, followed by a later, more effortful inferential process involving reasoning about the speaker's mental state (Kronmüller & Guerra, 2020). Other work indicates that listeners rapidly adapt to speaker-specific intonational patterns when exposed to unconventional usage (Roettger & Rimland, 2020). Furthermore, Yoon and Brown-Schmidt (2014, 2019) showed that listeners can adapt their interpretations of disfluencies in a speaker-specific manner, adjusting their expectations about upcoming referents based on each speaker's characteristic patterns of disfluency. These findings suggest

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

that listeners possess sophisticated mechanisms for tracking speaker-specific linguistic behaviors. However, the extent to which these mechanisms apply to more standard pragmatic inferences, such as MEI, in multiparty conversations, remains an open question, which we address in the present research.

The strength and pervasiveness of MEI as a cognitive strategy raise questions about its fundamental nature. Is MEI a flexible inference that can be readily modulated according to speaker characteristics, or does it function as a more rigid, default reasoning mechanism? Three views have been proposed. The first conceptualizes MEI as a basic inductive bias operating as a cognitive constraint specific to word learning (Markman, 1992). The second frames it as a learned regularity about language structure that emerges from experience with stable word-referent mappings (Frank et al., 2009). The third account views MEI as arising from reasoning about a speaker's communicative intentions and knowledge, suggesting that it is an inference that should be highly amenable to contextual modulation (Diesendruck & Markson, 2001). These accounts generate distinct predictions regarding MEI's sensitivity to speaker characteristics. The first view predicts minimal or no adaptation to speaker consistency, as MEI would operate as a relatively fixed bias regardless of the speaker. The learned regularity account suggests that MEI might adapt based on accumulated experience with different speaker types, but potentially more gradually. The intention-based inference view predicts more flexible and potentially rapid adaptation based on inferences about speaker reliability. Recent empirical evidence provides mixed support for these perspectives. For instance, studies show that MEI is sensitive to prosodic focus in 2-year-olds (Brody et al., 2024) and to speakers' membership in a social group (Weatherhead & White, 2021), suggesting flexibility. Yet other findings indicate that MEI may rely more on assumptions about linguistic conventions than on complex reasoning about speakers' epistemic states (Srinivasan et al., 2019). Additionally, bilingual children often demonstrate weaker MEI than their monolingual peers (Davidson et al., 1997), supporting the view that linguistic experience shapes this inference. These varying perspectives and findings lead to divergent predictions about whether and how MEI might be affected by speaker-specific characteristics, especially in multiparty contexts where multiple speakers with different referential patterns must be tracked simultaneously.

The concept of referential consistency is particularly relevant to understanding the potential modulation of MEI. In conversations, speakers tend to establish 'conceptual pacts' with listeners (Brennan & Clark, 1996) or 'referential precedents' (Barr & Keysar, 2002) by using consistent expressions to refer to the same objects. These regularities create expectations that influence subsequent interpretations. When a speaker violates their own referential precedent by using a new term for a previously named object, listeners experience comprehension difficulties, manifested as delayed object selection and eye-movement disruptions (Metzing & Brennan, 2003). However, these partner-specific difficulties typically emerge relatively late in processing, suggesting that they may not influence initial referential interpretations through MEI (Kronmüller & Barr, 2007; for a review, see Kronmüller & Barr, 2015). This observation raises the possibility that MEI, which appears to operate rapidly in referential processing, might be relatively immune to speaker-specific modulation – a hypothesis that aligns with a view of MEI as a default reasoning strategy or a learned regularity rather than a flexible inference on speakers' intentions and knowledge.

To test whether and how MEI might be modulated by speaker characteristics in multiparty conversation, we focus specifically on the dimension of referential

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

consistency. Our experimental paradigm examines how listeners process novel expressions from speakers with different histories of referential behavior – one who consistently uses the same expressions for the same objects across mentions (referentially consistent) and another who frequently introduces new expressions for previously named objects (referentially inconsistent). This multiparty design introduces a significant cognitive challenge for listeners, who must not only track referential patterns but also associate these patterns with specific speakers and potentially apply different interpretive strategies based on speaker identity. This increased cognitive load offers a particularly stringent test of MEI adaptation, as it requires listeners to maintain distinct speaker models while simultaneously engaging in real-time referential processing. Following established methodological approaches in referential communication research (e.g., Cooper, 1974; Tanenhaus et al., 1995), we employ eye-tracking measures to capture the real-time dynamics of referential processing, allowing us to distinguish early from late effects and to determine whether any modulation of MEI is speaker-specific or instead reflects a more context-general adaptation, whereby individual speaker characteristics jointly shape the conversational context without being tracked independently.

Having established the theoretical foundations for investigating speaker-specific influences on MEI, we now turn to the empirical evidence on speaker adaptation in language processing more broadly. This review will provide essential background for interpreting our findings concerning whether and how MEI is modulated in multiparty conversational settings.

1.1. Speaker-specific adaptations and context sensitivity in linguistic processing

A substantial body of research demonstrates that listeners rapidly adapt to various aspects of speaker-specific characteristics during language comprehension. These adaptations encompass multiple dimensions of speaker variability, including accent, speaking rate, emotional expression and voice quality (for a review, see Wu & Cai, 2026). For instance, listeners process speech differently based on accent familiarity, with comprehension improving for familiar accents (Anderson-Hsieh & Koehler, 1988; Bradlow & Pisoni, 1999). Similarly, neurophysiological evidence reveals that listeners integrate speaker identity with message content within 200–500 ms, showing distinct brain responses to speaker-message mismatches, such as hearing a child's voice saying 'like to go to pubs at night to drink and relax' (Van Berkum et al., 2008; Wu et al., 2025). These rapid adaptations suggest sophisticated mechanisms for tracking speaker characteristics and adjusting interpretive processes accordingly.

Particularly relevant to pragmatic inference is how listeners adapt to speaker reliability and consistency. Gardner et al. (2021) demonstrated that online interpretations of scalar adjectives are influenced by a speaker's perceived reliability, with listeners showing different patterns of anticipatory eye movements based on the speaker's history of adjective use. Even young children show sensitivity to speaker reliability, modulating contrastive inferences during online language comprehension based on previous speaker behavior (Ju et al., 2023). These findings collectively suggest that listeners maintain detailed models of individual speakers and adjust their interpretative processes based on speaker-specific linguistic patterns. The question that emerges is whether MEI – a seemingly automatic inference that guides referential interpretation – is subject to similar speaker-specific modulation, particularly in the

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

cognitively demanding context of multiparty conversation where listeners must track multiple speakers simultaneously.

While the evidence regarding speaker-specific adaptations presents a complex picture, research on context sensitivity provides clearer indications that pragmatic processing is, indeed, subject to broader contextual modulation. Listeners rapidly adapt their interpretative strategies based on various contextual factors, including the physical environment, discourse history and general conversational norms (Knoeferle, 2019; Yoon & Brown-Schmidt, 2019). This adaptability extends to pragmatic inferences in particular. For instance, Bosker et al. (2019) found that listeners adjust their interpretations of disfluencies based on the statistical regularities in how disfluencies pattern with subsequent referents in the discourse context, independent of specific speaker models.

The sensitivity of pragmatic inferences to broader contextual factors raises the possibility that MEI might similarly adapt to the general characteristics of the conversational environment, even if not to speaker-specific patterns. In multiparty conversations, where multiple speakers contribute to establishing communicative patterns and roles, listeners might form generalized expectations about referential practices based on the overall conversational context rather than maintaining distinct expectations for each individual speaker. This would represent a form of context-general adaptation that is more economical in terms of cognitive resources while still allowing for some flexibility in interpretation.

The distinction between speaker-specific and context-general adaptation is particularly relevant for understanding MEI in multiparty settings. If MEI operates as a default reasoning strategy or a learned regularity rather than a highly flexible pragmatic inference, we might expect limited speaker-specific modulation but potentially more substantial adaptation to the general referential consistency in the conversational context. This would support a view of MEI as a robust heuristic that provides an initial interpretative framework, which may be adjusted based on accumulated evidence about the reliability of this heuristic in the current conversation but not necessarily calibrated separately for each individual speaker.

2. The present study

Building on the theoretical considerations and empirical evidence outlined above, the present research investigates whether and how MEI is modulated by speaker characteristics in multiparty conversation. Specifically, we manipulated the referential consistency of speakers to determine: (1) whether MEI operates as a default reasoning strategy regardless of speaker and general conversational characteristics, (2) whether listeners adapt their application of MEI in a speaker-specific manner when one speaker is referentially consistent and another is inconsistent, or (3) whether listeners instead adapt to the general conversational context rather than to individual speakers.

To assess whether listeners adapt to individual speakers in a speaker-specific manner, we designed two experimental conditions. In the control condition, neither speaker used referentially inconsistent expressions, thereby establishing a context in which both were referentially consistent. In the experimental condition, however, one speaker consistently adhered to their referential precedents, while the other regularly violated them. Following our rationale, speaker-specific adaptation would manifest as reduced reliance on MEI when interacting with a referentially inconsistent speaker

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

relative to a consistent one, while the control condition served as a baseline for MEIs in the absence of referential inconsistency.

A central question of our study concerns the time course of this modulation. To examine this, we employed a visual-world eye-tracking paradigm in which participants followed instructions from two speakers while their eye movements were recorded, providing a real-time index of their interpretive processes. This methodology allows us to distinguish between different forms of adaptation: no adaptation (suggesting MEI is a rigid heuristic), speaker-specific adaptation (indicating sophisticated tracking of individual speaker patterns), or context-general adaptation (suggesting a more holistic adjustment to the conversational environment). By examining both behavioral choices and temporal dynamics of eye movements, we can determine whether any adaptations occur during initial processing or emerge later as part of more deliberate processes.

We conducted two experiments using this approach. Experiment 1 utilized a between-subjects design where participants interacted with either two referentially consistent speakers or one consistent and one inconsistent speaker. Experiment 2 employed a within-subjects design, in which all participants experienced both a control phase with consistent speakers and an experimental phase including one consistent and one inconsistent speaker. The design incorporated greater ecological validity through face-to-face interaction, included a manipulation check, increased the number of instances of inconsistent reference by two (a 20% increase), and added two distractor objects to the visual display. In addition, Experiment 2 was conducted in Spanish and tested a larger sample. The findings from these experiments – particularly the strong persistence of MEI across conditions and the absence of speaker-specific adaptation – provide valuable insights into the nature of MEI and its role in referential communication.

3. Experiment 1

In experiment 1, participants played a referential communication game that consisted of choosing pictures of unconventional objects – for which there is no English name – presented on a computer screen following the directions of two confederate speakers. The speakers' instructions were pre-recorded; however, participants believed that they were listening to live utterances produced by two speakers who were in a contiguous room, each communicating through a computer with microphones and headphones attached.

Participants were assigned to one of two groups. In the experimental group, one of the speakers – the referentially consistent speaker – used the same expression to refer to the same object throughout the experiment (Speaker A). In contrast, the other speaker – the referentially inconsistent speaker – would sometimes change the expression he previously used to name that same object (Speaker B). For example, the referentially inconsistent would call a strange object the 'futuristic building' and then, on a subsequent mention, would call it the 'window blinds'. In contrast, a referentially consistent speaker would always use the same expression to refer again to the same object. In the control conditions, both speakers (Speaker A and Speaker B) were referentially consistent. Participants were tested in the *test trials* where an object already named, the *named* object, and an object not yet named, the *unnamed* object, were presented on the screen and a speaker uttered a new expression not heard before but that could be mapped to any of the objects.

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

3.1. Methods

3.1.1. Participants

Thirty-two undergraduate students from the University of California, Riverside, participated in this study in exchange for credit in an introductory course in psychology. All were native English speakers. Eighteen of the participants were female. Three additional participants were excluded because of software failure.

3.1.2. Design

Experiment 1 had a 2 × 2 mixed-factorial design with *Group* (Experimental/Control) administered between subjects and *Speaker Identity* (Speaker A/Speaker B) administered within participants. In the experimental group, Speaker A was a referentially consistent speaker, following their own referential precedents, and Speaker B was a referentially inconsistent speaker. In the control group, both Speaker A and Speaker B were referentially consistent throughout the experiment.

3.1.3. Materials

The experiment included two types of item sets: experimental sets (16 in total) and manipulation sets (20 in total). Each experimental item set was built from a group of five black-and-white images depicting unfamiliar objects with no conventional names. Two of these objects shared physical characteristics and could therefore plausibly be labeled using the same expression; these were the *test objects*. (A complete list of test objects is provided in Supplementary Appendix 1, and a summary of the terminology used is provided in Supplementary Appendix 3.) The manipulation sets consisted of seven black-and-white images of similarly unconventional objects. No two objects within a manipulation set could be labeled with the same expression.

A total of 296 pre-recorded sound files were used – six for each experimental set and ten for each manipulation set. The expressions used in these recordings were constructed to uniquely identify the intended object in most contexts (i.e., when no visually similar object was present in the referential scene).

The experimental items consisted of three trials: a *presentation* trial, a *test* trial and one *filler* trial (see Figure 1, upper panel). The presentation trial always preceded the test trial. In half of the items, the filler trial occurred between the presentation and test trials; in the other half, it was placed either at the beginning or at the end of the item. In the presentation trial, one of the test objects was presented as the target alongside a distractor object. In the test trial, the two test objects were shown together. Two test expressions were created for each experimental item, each capable of referring to either test object. One expression was used in the presentation trial and the other in the test trial. Feedback was only given during the presentation trial.

Each manipulation item consisted of four trials: one *first-mention* trial, one *second-mention* trial, and two *filler* trials (see Figure 1, lower panel). The first-mention trial always preceded the second-mention trial, with one filler trial in between. The remaining filler trial was randomly placed either at the beginning or at the end of the item. The manipulation sets were further divided into two categories: *break* items and *maintain* items. For Speaker A in the experimental condition and both speakers, A and B, in the control condition, the same expression was used to refer to the same object across the first and second mention trials. In contrast, for

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

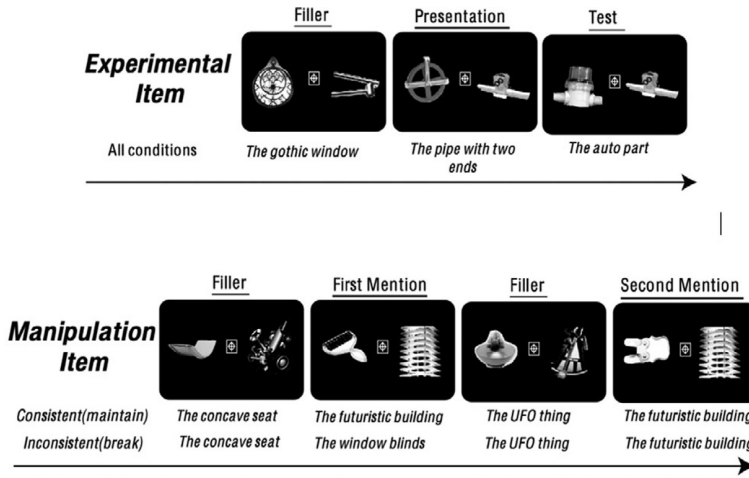


Figure 1. Upper panel. Example of an experimental item. In the presentation trial, the speaker uses an expression that could be easily mapped onto one of the objects. In the test trial, the previously named object appears again along a similar unnamed object. The speaker uses a new referring expression that could be mapped onto both objects. Lower panel. Example of a manipulation item. The referentially inconsistent speaker (Speaker B in the experimental group) would use two expressions to refer to the same object in the first and second mentions. The referentially consistent speakers (Speaker A in the experimental and control group and Speaker B in the control group) use the same expression.

Speaker B in the experimental condition, two different expressions were used to refer to the same object in the first and second mentions. Feedback was provided for both the first and second mention trials.

3.1.4. Procedure

Participants scheduled appointments to take part in the experiment. Two confederate speakers were instructed to arrive at the laboratory at the time of each participant's appointment and to pretend they were also naïve participants. The researcher provided a general overview of the procedure and simulated a random assignment of roles by asking the participant and the confederates to each draw one of three cards. All three cards were labeled 'matcher', but the confederates had been instructed to claim that their cards read 'speaker'.

Following the role assignment, the researcher gave instructions using a computer-based presentation. Participants were told that the goal of the experiment was to understand how people communicate referentially about objects in the absence of pointing or gaze. To that end, the speakers would be located in a different room from the matcher and would communicate via microphones and loudspeakers connected to the computers used during the task. After the instructions, the confederate speakers were asked to sit at their assigned computers, put on headphones, and wait for further instructions from the experimenter once he had joined the matcher in the adjacent room.

The matcher was then seated in a separate room, positioned in front of a computer, and asked to read a brief reminder of the task. Next, the eye-tracking camera was calibrated. Before the experiment began, the researcher simulated a brief conversation with the speakers. In reality, however, this interaction involved only the playback

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

of pre-recorded audio files. During the simulated exchange, the experimenter reminded the speakers of a few details and asked whether they were ready to begin, to which they replied affirmatively. This marked the start of the experimental session.

The matcher's task was to select one of two objects displayed on the screen by clicking with the computer mouse, following the verbal instructions of one of the speakers. Although participants believed that they were listening to live speech, they were actually hearing pre-recorded utterances delivered through individual sound files.

To introduce the manipulation, the first eight items presented were manipulation items – four produced by each speaker. After these initial trials, the experimental items and the remaining manipulation items were presented in a randomized order. Participants' picture selections and eye movements were recorded during the test trials as they followed the speakers' instructions.

3.1.5. Data analysis

Participants' final choices were analyzed using mixed-effects logistic regression, with subjects and items included as random effects. The variable Group was treated as a between-subjects factor and Speaker Identity as a within-subjects factor. Effect sizes, confidence intervals and observed power were calculated using Monte Carlo simulations based on the fitted mixed-effects model. For interpretability, we report standardized effect sizes by converting the beta coefficients for each fixed factor and interaction to Cohen's d (Chinn, 2000). Power was defined as the proportion of iterations in which the parameter of interest reached significance.

For the analysis of eye movements, we aimed to identify time windows in which looks to the named versus unnamed objects differed across conditions. To this end, we conducted a cluster-based permutation analysis. Eye-tracking data were collected at a sampling rate of 60 Hz and initially segmented into 300 ms bins (approximately 20 frames per bin). These were then linearly interpolated to obtain 50 ms bins. To reduce potential noise, a LOESS smoothing function with a span of 1.6 was applied to each data point. The data were subsequently aggregated by subject and by items for two separate analyses. To construct a null distribution, we generated 10,000 permuted datasets by randomly shuffling condition labels, while preserving the structure of the design – that is, one within-subject factor (Speaker Identity) and one between-subjects factor (Group). For each bin in each permuted dataset, we fitted a linear regression model and recorded the p-values and t-statistic associated with each term (intercept, main effects and interaction). Clusters were identified when two or more consecutive bins showed p-values equal to or below 0.05. For each such cluster, we computed a *cluster mass statistic* by summing the corresponding t-values and retained the largest mass from each permutation to create the null distribution. The same procedure was then applied to the real data, and observed cluster masses were compared against the null distribution to determine statistical significance. This analysis allowed us to assess effects of mutual exclusivity inference (MEI) by testing the intercept, main effects and interactions. (For a similar approach, see Barr et al., 2014; Kronmüller et al., 2017.)

Statistical analyses were performed in R (version 4.5.0). Mixed-effects logistic regressions were implemented with the lme4 package (Bates et al., 2015). Power analyses were carried out using simr (Green & MacLeod, 2016). Data wrangling and

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

visualization were conducted using functions from tidyverse (Wickham et al., 2019). (All scripts and data can be found at https://osf.io/dv4ts/.)

3.2. Results

3.2.1. Choice

Table 1 shows the mean percentage of unnamed object selection and the marginal means across conditions. Participants selected the unnamed object M = 71.5%, compared to M = 28.5% selection of the named object (β_0 = 1.036; Z = 7.174; p < 0.0001). In the experimental group, they selected the unnamed object M = 60.2% of the time compared to M = 82.8% in the control group (β_{group} = 1.20; Z = 4.810; p < 0.0001). No statistically reliable difference was found between the Speaker conditions nor the interaction. Table 2 shows the results from the mixed-effects logistic regression. The intercept tests that the new object is selected more than chance, which corresponds to 50%.

To examine the statistically null effects of Speaker Identity and its interaction with Group, we evaluated the effect sizes (Figure 2, top right panel) and observed power associated with each parameter (Table 2). For Speaker Identity, the standardized effect size was essentially zero (d = −0.006, 95% CI [−0.219, 0.223]). The interaction term also showed a small effect size (d = 0.139, 95% CI [−0.325, 0.651]). With the present sample size and design, the smallest detectable effect size for the interaction with 80% power is d = 0.66. Additionally, as shown in Table 2, the observed power for these effects was low, 0.041 for Speaker Identity and 0.087 for the interaction.

The pattern of choices for each individual across conditions is shown in the bottom panel of Figure 2. The proportion of unnamed object selections for Speaker A is plotted on the x-axis and the corresponding proportion for Speaker B on the y-axis. Each number in the graph indicates the number of participants exhibiting the same

Table 1. Cell and marginal means of selection of the unnamed alternative

	Group		
Speaker	Control	Experimental	Marginal
A	82%	61.7%	71.9%
B	83.6%	58.6%	71.2%
Marginal	82.8%	60.2%	

Table 2. Mixed-effects logistic regression on choice

Effect	Beta [95% CI]	Standard Error	z	p	Power [95% CI]
Intercept	1.036 [0.829 1.254]	0.144	7.174	< 0.001	1 [0.996 1]
Group	1.207 [0.821 1.641]	0.251	4.810	< 0.001	0.997 [0.996 0.999]
Speaker	−0.012 [−0.397 0.404]	0.212	−0.055	0.956	0.041 [0.030 0.055]
Group × Speaker	0.252 [−0.589 1.181]	0.425	0.592	0.554	0.087 [0.073. 0.106]

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

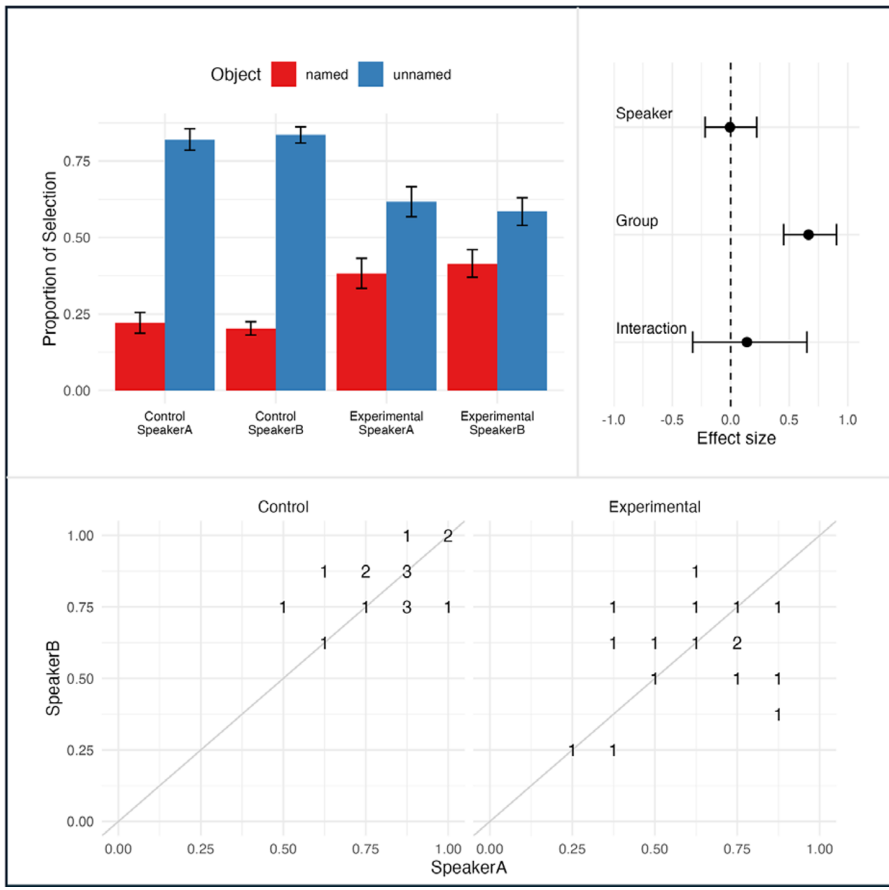


Figure 2. Results from the choice task. Top left panel: Mean proportion of object selections across conditions. Error bars represent the standard error of the mean, computed by bootstrapping across participants. Top right panel: Cohen's d adaptation derived from logistic regression, computed by bootstrapping across participants. Bottom panel: Individual patterns of unnamed object selection across conditions. The x-axis shows the proportion of unnamed selections for Speaker A and the y-axis for Speaker B; numbers indicate the number of participants exhibiting each response pattern. The gray diagonal represents indifference between speakers. Distance from the diagonal reflects speaker-specific adaptation: modulation of MEI based on speaker consistency is expected to fall below the diagonal in the experimental condition and remain near it in the control condition, whereas a context-general adaptation would shift responses toward the lower-left corner, indicating reduced selection of the unnamed alternative.

response pattern. The gray diagonal line represents indifference – that is, patterns of choice that do not differentiate between Speaker A and Speaker B, regardless of the overall preference for unnamed objects. Points above the diagonal reflect a pattern in which participants show a greater preference for the unnamed object for Speaker B but a preference for the named object for Speaker A, whereas the opposite pattern falls below the line. A greater distance from the diagonal indicates stronger speaker-specific adaptation. Accordingly, modulation of MEI based on speaker consistency should fall below the indifference line in the experimental condition and remain close to the indifference line in the control condition. A context-general adaptation would

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

shift responses toward the lower-left corner of the graph in the experimental condition, reflecting an overall reduction in selection of the unnamed alternative.

Descriptively, in the control condition, most participants selected the unnamed alternative independently of the speaker. These participants form a compact cluster in the upper-right region of the graph, close to the indifference line. In contrast, the experimental condition shows a less homogeneous pattern of responses. Most participants fall between the right corner and the center of the graph, indicating a reduced preference for the unnamed alternative. As in the control condition, most participants remain close to the indifference line, suggesting no systematic distinction between the consistent and inconsistent speaker, but more variable than the control condition.

3.2.2. Eye-movements *analysis*

Figure 3 left panel shows the proportion of looks to each object, time-locked to the onset of the referring expression. As shown, looks to the unnamed object steadily increase from the beginning of the observation window and continue rising throughout the time course. In contrast, looks to the named object decrease early in the window across all conditions. However, in the experimental group, this trend reverses around

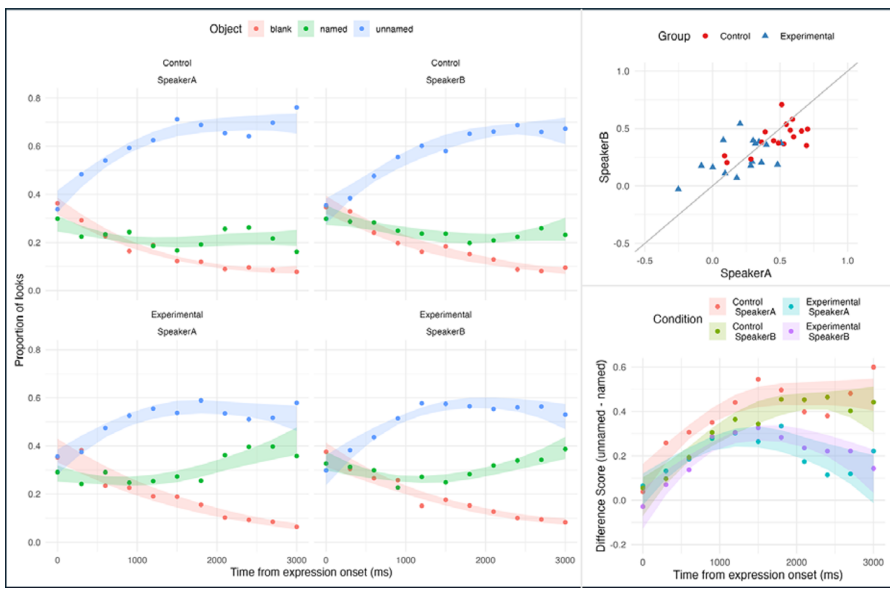


Figure 3. Eye-tracking results. Left panel: Proportion of looks to each object across conditions. Shaded error bands represent the standard error of the mean, computed by bootstrapping across participants and smoothed using a LOESS function; points indicate the raw (non-smoothed) data. Top right panel: Individual differences in difference scores, with Speaker A on the x-axis and Speaker B on the y-axis. Red points represent the control condition and blue triangles the experimental condition; the gray diagonal indicates indifference. Bottom right panel: Difference score (unnamed object − named object) from expression onset across conditions. Positive values indicate a preference for the unnamed object. Shaded error bands represent the standard error of the mean, computed by bootstrapping across participants; points indicate the raw (non-smoothed) data.

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

1000 ms, with looks to the mentioned object increasing and remaining elevated until the end of the time window.

To test these differences statistically, we conducted a cluster-based permutation analysis on the difference between looks to the unnamed and named objects (see Figure 3, bottom right panel). To identify the time window during which looks to the unnamed object diverged from looks to the named one, we searched for clusters defined by a statistically reliable intercept – testing whether the difference was significantly greater than zero. This analysis revealed a reliable cluster from 150 ms post-onset to the end of the time window at 3000 ms ($p < 0.0001$). Analysis by item revealed a reliable cluster ($p < 0.0001$) from 100 to 3000 ms.

When examining differences across conditions, we found a significant cluster only for the main effect of Group, spanning from 650 to 3000 ms ($p < .001$). Analysis by item yield a reliable cluster ($p = 0.004$) from 600 to 3000 ms. No reliable clusters were found for the main effect of Speaker or for the Group × Speaker interaction.

Individual responses show a general tendency to look at the unnamed object regardless of speaker, as reflected in positive difference scores for most participants clustered near the indifference line. In the experimental condition, responses shift toward the lower-left corner of the plot, indicating a modulation of MEI driven by the general conversational context rather than by speaker-specific adaptation, as there is no systematic deviation away from the indifference line.

3.3. Discussion

Participants engaged in mutual exclusivity inference (MEI), consistently selecting the unnamed object as the referent for a novel expression across all experimental conditions. However, in the experimental group, the presence of a referentially inconsistent speaker modulated this inference, resulting in a reduced selection of the unnamed object. Importantly, this modulation applied not only to the inconsistent speaker but also to the consistent one, suggesting that the adaptation was not speaker-specific. This pattern is consistent with a context-level adaptation to the presence of unreliable pragmatic cues in the conversation as a whole, rather than a speaker-specific adjustment. However, because this effect was observed in a between-subjects design with a relatively small sample, it remains to be confirmed whether it reflects a stable property of MEI or is specific to the conditions of Experiment 1. Experiment 2 was designed, in part, to address this question.

Real-time measures confirmed that MEI is a rapid inference, emerging approximately 150 ms after the onset of the referring expression across all conditions. The context-specific modulation appeared approximately 500 ms later. Crucially, no evidence was found for a transient speaker-specific adaptation at any point during the time window.

A descriptive analysis showed that most participants exhibited response patterns – across both choice behavior and eye movements – that were consistent with the aggregated data. However, some variability was observed in choice behavior.

Several caveats are worth noting. First, the lack of sensitivity to individual speakers may partly reflect the absence of real interaction. The use of pre-recorded speech and non-present speakers may have limited participants' ability to attribute and track speaker-specific behaviors – an issue highlighted in prior work emphasizing the role of interactive contexts in referential grounding (e.g., Brown-Schmidt, 2009).

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

Moreover, it is possible that participants did not construct fully independent speaker models, thereby limiting the generation of speaker-specific interpretations tailored to each interlocutor. Instead, participants may have formed a more global representation of the conversational group or 'team', interpreting referential expressions at this collective level rather than with respect to individual speakers.

Second, the referential context included only two objects, a design feature that may have encouraged anticipatory processing. In particular, the novelty of the unnamed object could have attracted visual attention even prior to linguistic input. Although this possibility cannot be entirely ruled out, the observation that MEI effects emerged at approximately 150 ms – a latency consistent with the minimal time required to program an eye movement – suggests that the inference was triggered by the linguistic input rather than by purely stimulus-driven attentional factors.

Third, the experiment had low observed power to detect speaker-specific modulation in the final choice analysis. This low power likely reflects a combination of a relatively small sample size and a genuinely small effect, as suggested by the mean selection rates for the unnamed alternative and individual variability in responses. This variability can be observed in the less homogenous pattern of responses among individual and the wide confidence intervals around the estimated effect size. These limitations were addressed in Experiment 2.

4. Experiment 2

Experiment 2 employed a design similar to that of Experiment 1, particularly regarding the nature of the manipulation involving referential consistency. Participants were exposed to both a referentially consistent and an inconsistent speaker. Unlike Experiment 1, however, the manipulation was implemented within subjects: all participants completed a baseline control phase prior to the manipulation, followed by the experimental phase. To enhance speaker-specific adaptation effects, participants interacted face-to-face with two confederate speakers. We include a manipulation check that quantitatively and qualitatively tests recognition of speaker characteristics. Additionally, the test screen included two filler objects: the *unseen* object never appeared before and the *seen* that appeared in previous trials but was never mentioned. This experiment was conducted entirely in Spanish with native Spanish-speaking participants.

4.1. Methods

4.1.1. Participants

Forty-four native Spanish-speaking undergraduate students from the Pontificia Universidad Católica de Chile participated in this study in exchange for 5.000 CLP. They all had normal or corrected-to-normal vision.

4.1.2. Design

The experiment employed a 2 × 2 within-subjects factorial design with the factors Phase (Control vs. Experimental) and Speaker Identity (Speaker A vs. Speaker B). The Control phase preceded the manipulation, while the Experimental phase followed it. During the manipulation trials, Speaker A was referentially consistent, whereas Speaker B was referentially inconsistent.

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

4.1.3. Apparatus

The experiment was conducted with Eye-Link 1000 remote eye-tracker at a frame rate of 500 Hz and programed with Experiment Builder Software (SR Research Ltd.).

4.1.4. Materials

The study included 24 test items, each consisting of three trials: a *grounding* trial, a *filler* trial and a *test* trial. In each trial, four images were presented on the display screen. These images depicted unusual objects or parts of objects that lacked conventional labels. Importantly, the objects used in each item set were unique and were not reused across different sets. Additionally, 40 manipulation items were created, each comprising two trials. As with test items, all trials in the manipulation items consisted of four images of unusual objects with no conventional name in Spanish.

Implementation of an experimental item. Figure 4 presents a schematic of an experimental item. Each item consisted of three trials. In the grounding trial, the speaker used a description to refer unambiguously to one of the objects. The filler trial was introduced to prevent participants from realizing the pattern of naming and to introduce, only visually, the *unnamed* object. Finally, in the *test* trials, from where the data analyzed was obtained, the same speaker uses a new expression that could be applied to two objects, the *named* object and the unnamed object (A complete list of test objects is provided in Supplementary Appendix 2). It also includes two distractor objects, an *unseen* distractor that never appeared before and a *seen* distractor that was seen but not mentioned (as the unnamed object). All experimental items were the same for all conditions, that is, the objects, position and expressions. What distinguished them was the position of the trial, either before the manipulation trials – in the control phase – or after the manipulation – in the experimental phase and the speaker who gave the instructions, Speaker A or Speaker B.

Implementation of a manipulation item. To manipulate the referential consistency of the speakers, we introduced 40 manipulation items in the middle of the session, in a way that 12 experimental items appeared before the manipulation trials and 12 after it. In the manipulation, the referentially consistent speaker (Speaker A) established a precedent in the first-mention trial and reused the same referring expression in the second-mention trial. In contrast, the referentially inconsistent speaker (Speaker B)



Figure 4. Example of a test item. The object in the upper right corner of the test screen is the named object. In the lower right, the unnamed. In the lower left, the unseen object, and in the upper left, the seen object. The Spanish words 'plano' (blueprint) and 'nave' (starship) apply to both named and unnamed objects.

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

also established a precedent in the first-mention trial but used a different expression in the second. Across the manipulation phase, Speaker A and Speaker B gave instructions for 20 manipulation trials each. Speaker B violated referential precedents on 12 occasions. In the top panel of Figure 5, an example of a manipulation item for the referentially consistent Speaker A is shown. As can be seen, the speaker establishes a precedent for one of the objects as a description that is used later in a second trial. In contrast, as can be seen in the lower panel, the referentially inconsistent, Speaker B, does not re-use the same expression, breaking its own precedent.

4.1.5. Procedure

Participants played a referential communication game with two confederate speakers, both co-present, in the context of a visual-world eye-tracking paradigm. Both speakers sat side-by-side looking at a screen where the scripted instructions were written. The participant sat in front in a way that could not see the speakers' screen nor could the speakers see the participant's screen. The four objects appeared on the participants' screen, and the task was to listen to one of the speakers' descriptions and select the target using a computer mouse. Unbeknownst to the participant, as said above, the speakers' screen displayed a pre-scripted instruction for each trial that they were to read aloud (the scripts and detailed instructions are available at https://osf.io/dv4ts/). Feedback was given in all but experimental test trials.

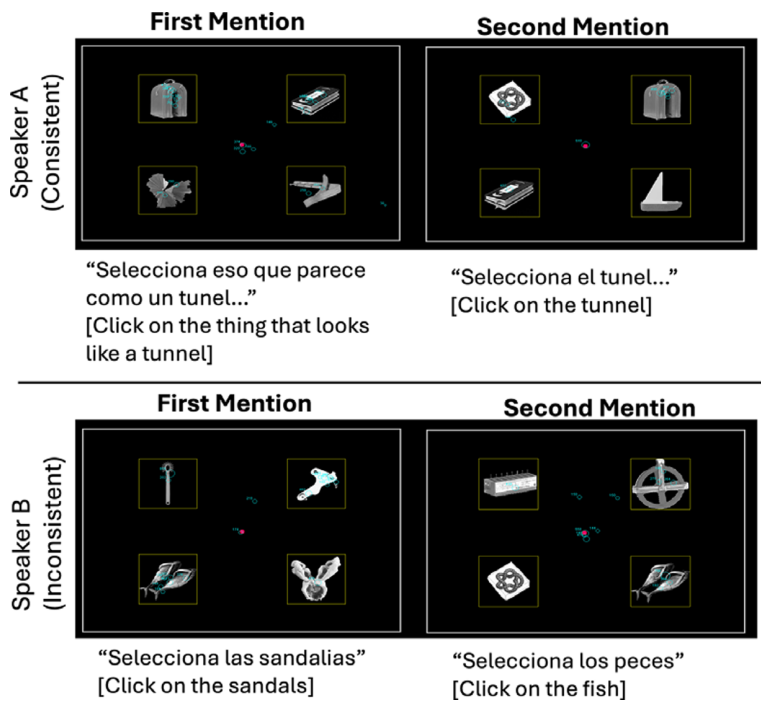


Figure 5. Example of manipulation trials. The top panel for Speaker A (referentially consistent) refers to the object in the upper left corner of the screen as the tunnel in the first mention and then again in the second-mention trial. The bottom panel for Speaker B (referentially inconsistent) refers to the object in the lower left corner of the screen in the first-mention trial as the sandals and then changes to fish in the second-mention trial.

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

We tracked participants' eye-movements with an Eye-Link 1000 remote eye-tracker at a frame rate of 500 Hz. To synchronize the eye data with the speakers' descriptions, we recorded the descriptions into sound files which were later tagged by a research assistant. We also recorded which object the matcher selected on each trial.

Cover story. We created a cover story that led participants to believe that the two speakers were actual participants rather than lab assistants so that they would not suspect they were aware of the main goals and hypotheses of the experiment. We adopted several strategies to make the speakers seem like real participants. First, one speaker arrived at the lab before and one after the participant. Second, we gave the instructions to all them three at the same time, where the confederates acted as if they were new to the task. And third, as we did in Experiment 1, we staged an apparent 'random' assignment of director/matcher roles by the experimenter announcing that they were holding three cards, two marked 'speaker' and the other 'listener', and having the participant and confederates each choose one without looking at them. The outcome was pre-determined because all three cards said 'listener', and the confederate always announced that they had drawn the role of 'speaker'.

Manipulation check. As a manipulation check, we asked participants to describe each speaker qualitatively, evaluate them and ask for reasons that justified the evaluation. First, they rated how clear and specific each speaker's instructions were on a scale from 1 to 7. For Speaker A, the mean was $M = 5.63$ ($SD = 1.27$) and for Speaker B, $M = 4.72$ ($SD = 1.42$). This difference was statistically reliable ($t_{83} = 3.119$, $p = 0.002$). Then, they indicated which speaker made the task 'easier' to complete. Thirty-two responded Speaker A, eleven Speaker B, and one participant did not respond.

We then ask for a description of each speaker and the reasons for choosing one of the Speakers as the one who was easier to complete the task. Twenty-six participants explicitly refer to one of the speakers as having either used the same expression to refer to the same object or changed a previous expression to refer again to the same object. Moreover, six of the eleven participants who responded that the task was easier with Speaker B (the inconsistent speaker) were among those who explicitly recognized the manipulation, but made the decision based on different criteria, such as that Speaker B was more creative, or that made more efforts to be understood, reasons that might point toward an implicit recognition of the manipulation (participants' responses are available at https://osf.io/dv4ts/.)

4.1.6. Data analysis

We implemented the same strategy as Experiment 1 for data analysis.

4.2. Results

4.2.1. Choice

Figure 6 top left panel shows the selection of each object across conditions. Participants selected the unnamed object $M = 63.5\%$, compared to $M = 33.2\%$ selection of the named object ($\beta_0 = 1.084$; $Z = 2.751$; $p = 0.006$). No main effect of phase nor main effect of speaker or interaction was found (See Table 4). As seen in Table 3, the selection rate for the unmentioned object was similar across all conditions.

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

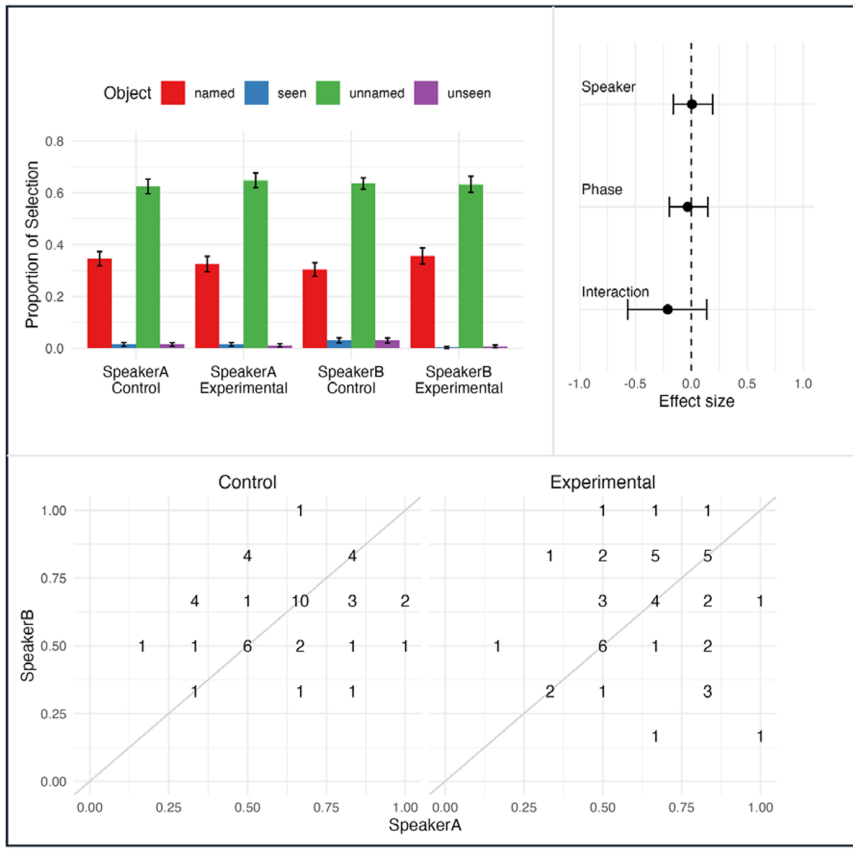


Figure 6. Results from the choice task for Experiment 2. Top left panel: Mean proportion of object selections across conditions. Error bars represent the standard error of the mean, computed by bootstrapping across participants. Top right panel: Cohen's d adaptation derived from logistic regression, computed by bootstrapping across participants. Bottom panel: Individual patterns of unnamed object selection across conditions. The x-axis shows the proportion of unnamed selections for Speaker A and the y-axis for Speaker B; numbers indicate the number of participants exhibiting each response pattern. The gray diagonal represents indifference between speakers. Distance from the diagonal reflects speaker-specific adaptation: modulation of MEI based on speaker consistency is expected to fall below the diagonal in the experimental condition and remain near it in the control condition, whereas a context-general adaptation would shift responses toward the lower-left corner, indicating reduced selection of the unnamed alternative.

Table 3. Cell and marginal means of selection of the unnamed alternative

	Phase		
Speaker	Control	Experimental	Marginal
A	62.5%	64.8%	63.6%
B	63.6%	63.2%	63.5%
Marginal	63.0%	64.0%	

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

To further examine the statistically null effects of Speaker Identity, Phase and their interaction, we evaluated the effect sizes (Figure 6 top right panel) and observed power (Table 4) associated with each parameter. For Speaker Identity, the standardized effect size was essentially zero ($d = -0.007$, 95% CI [−0.160, 0.191]). For Phase, the effect size was also very small ($d = -0.032$, 95% CI [−0.196, 0.148]). The interaction term also showed a small effect size ($d = -0.211$, 95% CI [−0.568, 0.138]). With our sample size and design for Experiment 2, the smallest detectable effect size for the interaction with 80% power is $d = -0.52$.

As shown in the bottom panel of Figure 6, individual response patterns exhibit variability in both preferences for one of the speakers and the general tendency toward MEI, with greater variability during the experimental phase than during the control phase. More participants are located in the lower-right corner of the graph in the experimental phase, indicating some degree of speaker sensitivity in their responses.

4.2.2. Eye-movements data

Figure 7 left panel shows the proportion of looks to each object, time-locked to the onset of the referring expression. Up to approximately 1000 ms, looks to both the named and unnamed objects increase in parallel, exceeding looks to the two distractor objects. After this point, participants begin to shift their gaze away from the named object and increasingly fixate on the unnamed object through the remainder of the observation window – consistent with mutual exclusivity inference (MEI).

We look for clusters where the difference score diverged from zero (Figure 7 bottom right panel). A positive value indicates a preference for the unnamed object and a negative for the named object. There is only one reliable cluster for the *intercept*, reflecting that overall, from 1200 to the end of the window at 3000 ms, there was a preference toward the unnamed object that was not modulated by our manipulations. This cluster was reliable by subjects ($p = 0.003$). Analysis by item revealed a reliable cluster ($p = 0.0005$) from 1250 to 3000 ms.

Finally, individual eye-movement response patterns show a largely homogeneous distribution across participants, with most responses clustered near the indifference line and shifted toward the upper-right corner. This pattern is consistent with a general tendency toward MEI (See Figure 7, top-right panel).

4.3. Discussion

We conducted Experiment 2 to further investigate the possibility of speaker-specific effects. In this version, we created a fully interactive setting in a fully within-subjects

Table 4. Mixed-effects logistic regression on choice

Effect	Beta [95% CI]	Standard error	*z*	*p*	Power [95% CI]
Intercept	1.084 [0.836, 1.24]	0.394	2.751	< 0.01	0.792 [0.766, 0.817]
Phase	−0.058 [−0.355, 0.269]	0.164	−0.354	0.723	0.062 [0.048, 0.079]
Speaker	0.0122 [−0.29, 0.347]	0.163	0.073	0.942	0.072 [0.057, 0.09]
Phase × Speaker	−0.382 [−1.03, 0.251]	0.327	−1.168	0.243	0.225 [0.199, 0.252]

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

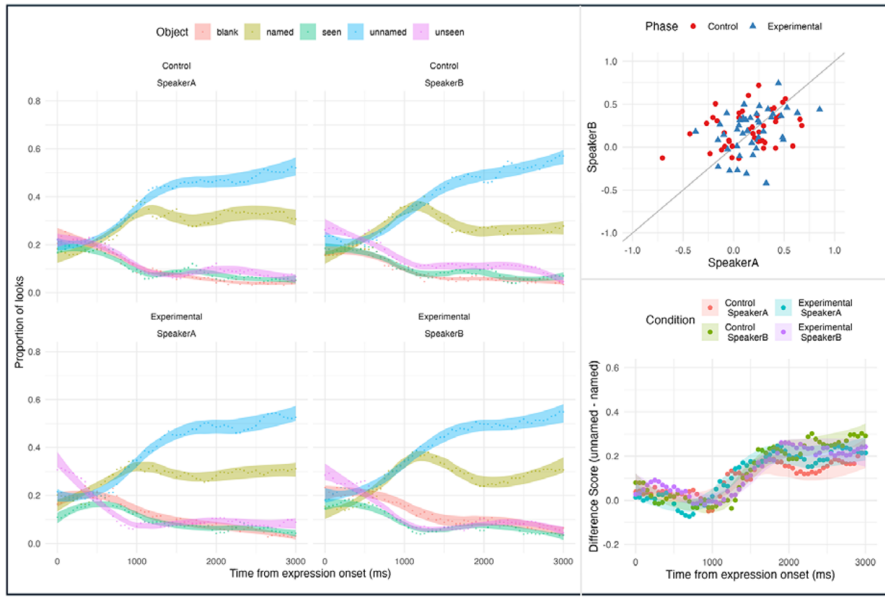


Figure 7. Eye-tracking results for Experiment 2. Left panel: Proportion of looks to each object across conditions. Shaded error bands represent the standard error of the mean, computed by bootstrapping across participants and smoothed using a LOESS function; points indicate the raw (non-smoothed) data. Top right panel: Individual differences in difference scores, with Speaker A on the x-axis and Speaker B on the y-axis. Red points represent the control condition and blue triangles the experimental condition; the gray diagonal indicates indifference. Bottom right panel: Difference score (unnamed object − named object) from expression onset across conditions. Positive values indicate a preference for the unnamed object. Shaded error bands represent the standard error of the mean, computed by bootstrapping across participants; points indicate the raw (non-smoothed) data.

design with a larger sample. To minimize anticipatory looks toward the unnamed object, we also included two distractor objects – one novel and one previously seen but unnamed. The results showed a robust effect of mutual exclusivity inference (MEI), confirming that participants reliably inferred that a novel expression referred to a previously unnamed object. However, even in this interactive context, no speaker-specific modulation of MEI was observed. Moreover, the contextual adaptation observed in Experiment 1 did not replicate, as no significant difference was found between the experimental and control conditions (we elaborate further on this issue in the next section). Finally, the inclusion of additional objects appears to have increased competition between the named and unnamed targets, delaying MEI relative to Experiment 1 and ultimately reducing selection of the unnamed alternative compared to Experiment 1.

The absence of speaker sensitivity cannot be explained by confusion or a lack of encoding of speaker characteristics. Indeed, participants explicitly reported that the task was easier to conduct with Speaker A, and in most cases, they mentioned that the reason was the inconsistency in Speaker B's naming patterns or the consistency in Speaker A's patterns. Moreover, half of the participants who reported that the task was easier to complete with Speaker B explicitly recognized the manipulation. Related to power, the low observed power for the interaction in the selection tests may be

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

attributable to a genuinely small effect combined with variability in responses, with only a few participants showing evidence of speaker sensitivity.

(Although a subset of participants indicated that the inconsistent speaker made the task easier, we conducted all analyses restricting the sample to those who found the consistent speaker easier to work with. The pattern of results remained unchanged. Analyses are available at https://osf.io/dv4ts/.)

5. General discussion

This study investigated whether listeners adapt their pragmatic inferences based on speakers' referential consistency in multiparty conversations. By manipulating whether speakers consistently used the same expressions to refer to objects (referentially consistent) or sometimes changed their referring expressions (referentially inconsistent), we examined whether MEI – the inference that a novel term refers to a previously unnamed object – is modulated in a speaker-specific manner. Across two experiments, three key findings emerged. First, MEI proved to be a robust inferential strategy. Participants consistently mapped novel expressions onto unnamed objects, demonstrating MEI's fundamental role in resolving referential ambiguity. Second, MEI was not reliably modulated by speaker-specific characteristics. Although participants in both experiments were exposed to speakers with distinctly different patterns of referential consistency, and they were able to track and encode speaker characteristics, they did not adjust their MEI differentially based on speaker identity. This absence of speaker-specific adaptation was observed in Experiment 1 using pre-recorded stimuli and persisted in Experiment 2 despite the face-to-face interaction format and evidence from post-experiment questionnaires that participants explicitly recognized differences in speaker consistency. Third, we observed an adaptation to the general conversational context in Experiment 1. Participants in the experimental group (where one inconsistent speaker was present) showed an overall reduced tendency to apply MEI compared to the control group (where both speakers were consistent). However, this effect was not replicated in Experiment 2, which featured a larger sample, face-to-face interaction, and a within-subjects design in which all participants first experienced two consistent speakers. This non-replication admits at least two interpretations: the Experiment 1 result may reflect a genuine but fragile form of contextual adaptation that is suppressed under greater cognitive and interactional demands, or it may constitute a false positive linked to the between-subjects design or relatively small sample of that experiment. The present data do not allow us to adjudicate between these possibilities.

5.1. Theoretical contribution

Our findings support the conceptualization of MEI as a default reasoning strategy rather than a highly flexible inference. The robustness of MEI across conditions and its early emergence in eye-tracking data suggest it functions as an initial interpretive heuristic – a cognitive shortcut applied before more deliberate, context-sensitive processes engage. This aligns with accounts that frame MEI as a basic inductive bias (Markman, 1992) or learned regularity about language structure (Frank et al., 2009) rather than as a highly flexible pragmatic inference based on reasoning about speaker intentions and knowledge (Diesendruck & Markson, 2001).

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

The persistence of MEI despite explicit recognition of speaker inconsistency (in Experiment 2) suggests that this inference may be relatively impervious to top-down modulation by speaker-specific information. This finding contrasts with work demonstrating that listeners can adapt contrastive inferences (Grodner & Sedivy, 2011), scalar implicatures (Pogue et al., 2016) and disfluency processing (Yoon & Brown-Schmidt, 2014) in a speaker-specific manner. This divergence might reflect that different types of pragmatic inferences may vary in their susceptibility to speaker-specific modulation. MEI may be more deeply entrenched as a default interpretive strategy compared to other inferences, making it less amenable to rapid speaker-specific calibration. Alternatively, the cognitive demands imposed in multi-party conversation may exceed those in the dyadic settings used in most previous studies, explaining the apparent discrepancy in findings.

On the other hand, our findings are consistent with previous research showing that listeners adjust pragmatic inferences in response to the broader statistical properties of their linguistic environment (Bosker et al., 2019; Ryskin et al., 2019), even in the absence of sensitivity to speaker-specific information. In Experiment 1, listeners appeared to modulate MEI based on the general conversational context but not on individual speaker characteristics, although this context-level effect did not replicate in Experiment 2. If reliable, this pattern would be consistent with a hierarchical organization of adaptation processes, in which context-level adjustments precede or override speaker-specific adaptations, particularly under cognitive load. However, given the non-replication, this interpretation remains tentative and awaits confirmation under conditions that isolate context-level from design-level explanations. Managing multiple speakers in multiparty conversation imposes substantial cognitive demands, requiring listeners to monitor individual behaviors while maintaining real-time comprehension. Under such conditions, prioritizing broader, context-level adaptations may constitute an efficient strategy to manage processing resources.

Our findings inform ongoing debates on the temporal dynamics of pragmatic processing. In Experiment 1, MEI emerged rapidly, at approximately 150 ms after the onset of the referring expression, whereas context-driven modulation was observed at a later latency (~650 ms). This temporal separation, if robust, would align with dual-process frameworks of language comprehension in conversational contexts (Barr, 2008; Keysar et al., 2000; Kronmüller et al., 2017), which distinguish between an early, fast-acting inferential system and a subsequent, context-sensitive integration stage. It is important to note, however, that this dissociation is based solely on Experiment 1, since the context-driven component did not emerge in Experiment 2. Thus, while the early onset of MEI itself is well supported across both experiments, the two-stage temporal profile should be interpreted with caution.

However, the observed dynamics are also consistent with models of incremental pragmatic processing (Nieuwland et al., 2010; Rubio-Fernandez & Jara-Ettinger, 2020; Tanenhaus & Brown-Schmidt, 2008), which posit a continuous, cue-weighted updating mechanism wherein comprehenders flexibly integrate multiple information sources as they become available, without committing to strictly modular stages. Moreover, the delayed onset of MEI in Experiment 2 relative to Experiment 1 – by approximately 500 ms – may reflect increased integration demands. The greater amount of information available in the referential domain could have imposed higher processing demands, and the introduction of two additional objects could have contributed to this effect.

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

More generally, the absence of speaker-specific adaptation, even in the interactive setting of Experiment 2, suggests that integrating speaker-specific information may exceed available cognitive resources in multiparty conversation. This aligns with research on the cognitive demands of perspective-taking (Brown-Schmidt, 2009; Wardlow, 2013), which shows that maintaining and utilizing speaker-specific models is effortful and may be compromised under complex task conditions. Our findings suggest a potential limit to the granularity of speaker models that listeners can effectively integrate online in multiparty settings.

Finally, the differences between Experiment 1 and Experiment 2 require careful consideration. The failure to replicate context-general adaptation in Experiment 2 is open to multiple interpretations. One possibility is that it reflects an upper bound on adaptation under increasing cognitive and interactional demands: in Experiment 1, the communicative environment was comparatively simpler, with only two potential referents and reduced interactional complexity, and listeners may be better able to exploit broader statistical regularities under such conditions. However, alternative explanations cannot be ruled out. Experiment 1 effect may have been inflated by the between-subjects design, in which individual differences between groups could mimic a context-level adaptation. It is also possible that the fixed order of conditions in Experiment 2, in which all participants first experienced two consistent speakers, established strong expectations of reliability that the subsequent inconsistency was insufficient to override, a pattern that would be predicted by a rational Bayesian learner who weighs cumulative evidence (cf. Tauber et al., 2017). Disentangling these possibilities will require designs that independently manipulate cognitive demands, evidential strength and order of exposure. Consistent with this interpretation, recent evidence indicates that in non-interactive contexts listeners can track general statistical patterns of communicative behavior produced by a non-target speaker and use this information when interpreting utterances from a target speaker (Wu et al., 2025). Taken together, these findings suggest that context-general adaptation may be more readily expressed in simplified or less demanding communicative settings, whereas increased interactional complexity may constrain the deployment of such adaptations online.

5.2. Limitations and future research

Some limitations of the current study suggest avenues for future research. First, the cognitive load imposed by our multiparty design may have constrained listeners' ability to integrate speaker-specific information during online interpretation. Future studies could manipulate cognitive load directly or employ simpler designs, with a single speaker, to determine whether speaker-specific adaptation of MEI emerges under less demanding conditions.

Second, our manipulation of referential inconsistency was relatively subtle, involving changes in object descriptions rather than more overt violations of conversational norms. Stronger manipulations – such as a speaker who repeatedly introduces novel labels for common objects – may yield more pronounced effects on MEI.

Third, and closely related to this issue, the overall frequency of inconsistent references may have been too low to reliably prompt adaptation (Rivera-Vera et al., 2022). Future work could address this by systematically manipulating the ratio of consistent to inconsistent instances in order to identify potential thresholds for adaptive behavior. Moreover, such thresholds may vary across individuals. This

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

possibility could be explored using a Bayesian modeling framework that explicitly incorporates individual differences in sensitivity to inconsistency, allowing participant-specific parameters to capture variability in adaptation dynamics (Tauber et al., 2017).

Fourth, our definition of context in the present study is limited to the properties of a specific conversational interaction. Nevertheless, context may also be construed more broadly, encompassing population-level or socially defined characteristics. Future work could explore whether adaptation extends to such broader contextual dimensions. In addition, stronger evidence for this context-general adaptation would require designs involving a larger number of speakers or direct tests of generalization to new speakers or different conversational contexts.

Finally, individual differences in cognitive abilities like working memory, inhibitory control and perspective-taking might influence listeners' capacity to track and utilize speaker-specific information. Recent studies have shown that personality traits can influence the integration of contextual information during language interpretation online (Wu et al., 2025; Wu & Cai, 2025). Future studies could incorporate measures of these abilities to explore their relationship with pragmatic adaptation.

5.3. Broader implications

Beyond its theoretical significance, this research has practical implications for communication in multiparty contexts. Our findings suggest that communicative efficiency might be optimized by maintaining referential consistency across speakers rather than expecting listeners to adapt to individual speakers' idiosyncratic patterns. This has relevance for contexts ranging from classroom instruction to business meetings to clinical interactions, where multiple speakers often contribute to a shared discourse. For technological applications like dialog systems and virtual agents, our results suggest that implementing consistent referential practices across multiple artificial speakers might be more important than modeling individual speaker characteristics in multiparty human–computer interaction.

6. Conclusion

Inferences based on mutual exclusivity function as a robust default reasoning strategy in referential communication.

While listeners can adapt their application of MEI to the general conversational context, they appear unable to calibrate this adaptation in a speaker-specific manner in multiparty conversation. These findings advance our understanding of the cognitive mechanisms underlying pragmatic inference and the constraints on pragmatic adaptation in complex communicative environments. They suggest that MEI occupies a special status among pragmatic inferences – relatively resistant to speaker-specific modulation yet potentially sensitive to broader contextual factors. This reflects MEI's foundational role in referential communication, providing a stable interpretive framework that nonetheless exhibits some contextual flexibility.

Across two experiments employing different designs, languages and levels of interactional complexity, listeners did not calibrate MEI in a speaker-specific manner, even when they explicitly encoded differences in speaker consistency. Evidence for adaptation to the general conversational context was observed in Experiment

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

1 but did not replicate in Experiment 2, leaving the scope and conditions of such adaptation as an open question. These findings advance our understanding of the constraints on pragmatic adaptation in complex communicative environments. Rather than exhibiting a flexible sensitivity to speaker or contextual characteristics, MEI proved remarkably stable under the demanding conditions of multiparty conversation, suggesting that its primary function is to provide a reliable interpretive default precisely in situations where tracking multiple information sources may exceed available cognitive resources.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/langcog.2026.10086.

Acknowledgments. This research was supported by the Fondecyt-Chile Grant (1200655 and 11100226), and funding from ANID CIAE CIA250005 is gratefully acknowledged. We thank Dale Barr for his valuable input during the development of Experiment 1. This manuscript involved the use of OpenAI's ChatGPT (version GPT-4o) and Antropic's Claude (Opus 4.6) to improve the clarity and conciseness of writing. The tool was used solely for editing purposes; all content, structure and scientific interpretations were developed entirely by the authors. The authors reviewed and verified all suggestions made by the tool to ensure accuracy and integrity.

Ethics. The study design, experimental protocols and informed consents were approved by the University of California, Riverside, Ethics Committee for Experiment 1 and by the FONDECYT-CHILE Ethics Committee for Experiment 2.

References

Anderson-Hsieh, J., & Koehler, K. (1988). The effect of foreign accent and speaking rate on native speaker comprehension. *Language Learning*, 38, 561–613. https://doi.org/10.1111/J.1467-1770.1988.TB00167.X

Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition*, 109(1), 18–40. https://doi.org/10.1016/j.cognition.2008.07.005

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

Barr, D. J., & Keysar, B. (2002). Anchoring Comprehension in Linguistic Precedents. *Journal of Memory and Language*, 46(2), 391-418. https://doi.org/10.1006/jmla.2001.2815

Barr, D. J., Jackson, L., & Phillips, I. (2014). Using a voice to put a name to a face: The psycholinguistics of proper name comprehension. *Journal of Experimental Psychology: General*, 143(1), 404–413. https://doi.org/10.1037/a0031813

Bosker, H. R., van Os, M., Does, R., & van Bergen, G. (2019). Counting 'uhm's: How tracking the distribution of native and non-native disfluencies influences online language comprehension. *Journal of Memory and Language*, 106, 189–202. https://doi.org/10.1016/j.jml.2019.02.006

Bradlow, A. R., & Pisoni, D. B. (1999). Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors. *The Journal of the Acoustical Society of America*, 106(4), 2074–2085. https://doi.org/10.1121/1.427952

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(6), 1482–1493. https://doi.org/10.1037/0278-7393.22.6.1482

Brody, G., Pomiechowska, B., Csibra, G., & Gliga, T. (2024). Why do children think words are mutually exclusive? *Current Opinion in Psycholinguistics*, 8, 100–108. https://doi.org/10.1177/09567976241287732

Brown-Schmidt, S. (2009). The role of executive function in perspective taking during online language comprehension. *Psychonomic Bulletin & Review*, 16(5), 893–900. https://doi.org/10.3758/PBR.16.5.893

Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, 19(22), 3127–3131. https://doi.org/10.1002/1097-0258(20001130)19:22<3127::AID-SIM784>3.0.CO;2-M

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*. https://doi.org/10.1016/0010-0285(74)90005-X

Davidson, D., Jergovic, D., Imami, Z., & Theodos, V. (1997). Monolingual and bilingual children's use of the mutual exclusivity constraint. *Journal of Child Language*, 24(1), 3–24. https://doi.org/10.1017/S0305000996002917

Diesendruck, G., & Markson, L. (2001). Children's avoidance of lexical overlap: A pragmatic account. *Developmental Psychology*, 37(5), 630–641. https://doi.org/10.1037/0012-1649.37.5.630

Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2009). Using speakers' referential intentions to model early cross-situational word learning. *Psychological Science*, 20(5), 578–585. https://doi.org/10.1111/j.1467-9280.2009.02335.x

Gardner, B., Dix, S., Lawrence, R., Morgan, C., Sullivan, A., & Kurumada, C. (2021). Online pragmatic interpretations of scalar adjectives are affected by perceived speaker reliability. *PLoS One*, 16(2), e0245130. https://doi.org/10.1371/journal.pone.0245130

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. https://doi.org/10.1111/2041-210X.12504

Grodner, D., & Sedivy, J. C. (2011). The effect of speaker-specific information on pragmatic inferences. In N. Pearlmutter & E. Gibson (Eds.), *The processing and acquisition of reference* (pp. 239–272). MIT Press. https://doi.org/10.7551/mitpress/9780262015127.003.0010

Ju, N., Williams, N., Sedivy, J., Chambers, C. G., & Graham, S. A. (2023). Preschoolers modulate contrastive inferences during online language comprehension. *Child Development*, 94(5), 1319–1329. https://doi.org/10.1111/cdev.13925

Kachergis, G., Yu, C., & Shiffrin, R. M. (2012). An associative model of adaptive inference for learning word-referent mappings. *Psychonomic Bulletin & Review*, 19, 317–324. https://doi.org/10.3758/s13423-011-0194-6

Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: The role of mutual knowledge in comprehension. *Psychological science*, 11(1), 32–38. https://doi.org/10.1111/1467-9280.002

Knoeferle, P. (2019). Predicting (variability of) context effects in language comprehension. *Journal of Cultural Cognitive Science*, 3, 141–158. https://doi.org/10.1007/s41809-019-00025-5

Kronmüller, E., & Barr, D. J. (2007). Perspective-free pragmatics: Broken precedents and the recovery-from-preemption hypothesis. *Journal of Memory and Language*, 56(3), 436–455. https://doi.org/10.1016/j.jml.2006.05.002

Kronmüller, E., & Barr, D. J. (2015). Referential precedents in spoken language comprehension: A review and meta-analysis. *Journal of Memory and Language*, 83, 1–19. https://doi.org/10.1016/j.jml.2015.03.008

Kronmüller, E., & Guerra, E. (2020). Processing speaker-specific information in two stages during the interpretation of referential precedents. *Frontiers in Psychology*, 11, 552368. https://doi.org/10.3389/fpsyg.2020.552368

Kronmüller, E., Noveck, I., Rivera, N., Jaume-Guazzini, F., & Barr, D. (2017). The positive side of a negative reference: The delay between linguistic processing and common ground. *Royal Society Open Science*, 4(2), 160827. https://doi.org/10.1098/rsos.160827

Lewis, M., Cristiano, V., Lake, B. M., Kwan, T., & Frank, M. C. (2020). The role of developmental change and linguistic experience in the mutual exclusivity effect. *Cognition*, 198, 104191. https://doi.org/10.1016/j.cognition.2020.104191

Markman, E. M. (1992). Constraints on word learning: Speculations about their nature, origins, and domain specificity. In M. R. Gunnar & M. Maratsos (Eds.), *Modularity and constraints in language and cognition* (pp. 59–101). Erlbaum.

Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, 20(2), 121–157. https://doi.org/10.1016/0010-0285(88)90017-5

Metzing, C., & Brennan, S. E. (2003). When conceptual pacts are broken: Partner-specific effects on the comprehension of referring expressions. *Journal of Memory and Language*, 49(2), 201–213. https://doi.org/10.1016/S0749-596X(03)00028-7

Nieuwland, M. S., Ditman, T., & Kuperberg, G. R. (2010). On the incrementality of pragmatic processing: An ERP investigation of informativeness and pragmatic abilities. *Journal of Memory and Language*, 63(3), 324–346. https://doi.org/10.1016/j.jml.2010.06.005

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press

Pogue, A., Kurumada, C., & Tanenhaus, M. K. (2016). Talker-specific generalization of pragmatic inferences based on under- and over-informative prenominal adjective use. *Frontiers in Psychology*, 6, 2035. https://doi.org/10.3389/fpsyg.2015.02035

Rivera-Vera, N. A., Andringa, S., Kronmuller, E., Monaghan, P., & Rispens, J. E. (2022). The effect of speaker reliability on adult cross-situational word learning. *Glossa Psycholinguistics*, 1(1). https://doi.org/10.5070/G601113

Roettger, T. B., & Rimland, K. (2020). Listeners' adaptation to unreliable intonation is speaker-sensitive. *Cognition*, 204, 104372. https://doi.org/10.1016/j.cognition.2020.104372

Rubio-Fernandez, P., & Jara-Ettinger, J. (2020). Incrementality and efficiency shape pragmatics across languages. *Proceedings of the National Academy of Sciences*, 117(24), 13399–13404. https://doi.org/10.1073/pnas.1922067117

Ryskin, R., Kurumada, C., & Brown-Schmidt, S. (2019). Information integration in modulation of pragmatic inferences during online language comprehension. *Cognitive Science*, 43(8), e12769. https://doi.org/10.1111/cogs.12769

Srinivasan, M., Foushee, R., Bartnof, A., & Barner, D. (2019). Linguistic conventionality and the role of epistemic reasoning in children's mutual exclusivity inferences. *Cognition*, 193, 104036. https://doi.org/10.1016/j.cognition.2019.04.001

Tanenhaus, M. K., & Brown-Schmidt, S. (2008). Language processing in the natural world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 1105–1122. https://doi.org/10.1098/rstb.2007.2162

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632–1634. https://doi.org/10.1126/science.7777863

Tauber, S., Navarro, D. J., Perfors, A., & Steyvers, M. (2017). Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review*, 124(4), 410. https://doi.org/10.1037/rev0000052

Van Berkum, J. J. A., van den Brink, D., Tesink, C. M. J. Y., Kos, M., & Hagoort, P. (2008). The neural integration of speaker and message. *Journal of Cognitive Neuroscience*, 20(4), 580–591. https://doi.org/10.1162/jocn.2008.20054

Wardlow, L. (2013). Individual differences in speakers' perspective taking: The roles of executive control and working memory. *Psychonomic Bulletin & Review*, 20(4), 766–772. https://doi.org/10.3758/s13423-013-0396-1

Weatherhead, D., & White, K. S. (2021). Toddlers link social and speech variation during word learning. *Developmental Psychology*, 57(8), 1195–1209. https://doi.org/10.1037/dev0001032

Wickham, H, *et al.* (2019). Welcome to the Tidyverse. *The Journal of Open Source Software*, 4(43), 1686. https://doi.org/10.21105/joss.01686.

Wu, H., & Cai, Z. G. (2025). When a man says he is pregnant: Event-related potential evidence for a rational account of speaker-contextualized language comprehension. *Journal of Cognitive Neuroscience*, 38, 1–16. https://doi.org/10.1162/JOCN.a.102

Wu, H., & Cai, Z. G. (2026). Speaker effects in language comprehension : An integrative model of language and speaker processing. *Psychonomic Bulletin & ; Review*, 33(4). https://doi.org/10.3758/s13423-026-02896-6

Wu, H., Duan, X., & Cai, Z. G. (2024). Speaker demographics modulate listeners' neural correlates of spoken word processing. *Journal of Cognitive Neuroscience*, 36(10), 2208–2226. https://doi.org/10.1162/jocn_a_02225

Wu, H., Rao, X., & Cai, Z. G. (2025). Probabilistic adaptation of language comprehension for individual speakers: Evidence from neural oscillations. *Social Cognitive and Affective Neuroscience*, 20(1), 1–9. https://doi.org/10.1093/scan/nsaf085

Yoon, S. O., & Brown-Schmidt, S. (2014). Adjusting conceptual pacts in three-party conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(4), 919–937. https://doi.org/10.1037/a0036161

Yoon, S. O., & Brown-Schmidt, S. (2019). Contextual integration in multiparty audience design. *Cognitive Science*, 43(12), e12807. https://doi.org/10.1111/cogs.12807

Cite this article: Kronmüller, E., & Guerra, E. (2026). Adapting pragmatic interpretation in multiparty conversations: mutual exclusivity inferences across speaker and context, *Language and Cognition*, 18, e43, 1–27. https://doi.org/10.1017/langcog.2026.10086

https://doi.org/10.1017/langcog.2026.10086 Published online by Cambridge University Press