Acta Psychologica 255 (2025) 104908

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Differentiation and integration: The addressee perspective-taking strategy in three-party conversation

Xiaobei Zheng [a], Chao Sun [b,*]

[a] College of International Studies, Shenzhen University, Shenzhen, China
[b] School of Chinese as a Second Language, Peking University, Beijing, China

ARTICLE INFO

Keywords:
Perspective-taking
Common ground
Three-party conversation
Referential resolution
Mandarin

ABSTRACT

In conversations, people take the perspectives of others when engaging in referential understanding. Previous studies have primarily focused on dialogues between two participants. However, as the number of conversational partners increases, the perspective-taking strategies may change. This study specifically investigated a situation in which the addressee faces two speakers who take turns giving referential instructions. In Experiment 1, the perspectives of Speaker 1 and the addressee participants were consistent, while the perspectives of Speaker 2 and the addressee participants were inconsistent. In Experiment 2, the perspectives of both speakers were consistent but differed from the addressee's perspective. The results showed that, in Experiment 1, participants distinguished between the perspectives of the two speakers when interpreting noun reference, but no difference was found in Experiment 2. However, when comparing the results of Experiments 1 and 2, it was found that, despite the perspective of Speaker 2 remaining unchanged, participants in Experiment 1 were more egocentric than in Experiment 2 when interpreting Speaker 2's discourse. The pattern of strategic change was aligned with the interpretation of Speaker 1. This suggests that participants, to some extent, integrate the perspectives of both speakers. The results were further discussed based on the consideration of their partner's audience design strategies, use of a "Grounding by Proxy" strategy, or the calculation of a probabilistic weight of different perspectives.

1. Introduction

1.1. Common ground and perspective-taking

Dialogue serves as the primary means by which people exchange information. As speakers, individuals construct linguistic expressions to convey their intended messages. This process not only requires organizing language based on content but also necessitates accounting for contextual constraints. Similarly, listeners must interpret these expressions by considering the same constraints. Among these constraints, *common ground* is particularly critical. It refers to the shared knowledge between interlocutors in a conversation (Clark & Clark, 1977; Stalnaker, 2002). This knowledge is based on pre-existing information and dynamically evolves through interaction (Clark & Brennan, 1991; Clark & Wilkes-Gibbs, 1986). For example, during a conversation, if both individuals are aware that there is a specific movie playing, they might simplify their references to "the movie" rather than repeating its full title and subsequently use the pronoun "it" in their dialogue. Such reliance on common ground allows for more efficient and natural communication.

Common ground facilitates a variety of communicative processes, including the interpretation of figurative language (Gerrig & Horton, 2004; Gildea & Glucksberg, 1983; Ritchie, 2004), the understanding of a speaker's beliefs and attitudes (Harris & Potts, 2009; Kärkkäinen, 2006), and the resolution of referential domains in discourse (Clark & Marshall, 1978, 1981). The present study focuses specifically on how common ground contributes to referential resolution — that is, how interlocutors adjust the scope of lexical items based on their shared knowledge. Through joint attention and coordination within shared referential domains, interlocutors dynamically construct a shared cognitive framework. This process enables participants to align their understanding of words with contextual constraints, thereby making communication more efficient and fluid (Beun & Cremers, 1998; Jara-Ettinger & Rubio-Fernandez, 2022; Krauss & Fussell, 1991; Mozuraitis et al., 2015). This process is also referred to as *perspective-taking*, the ability to interpret information from another person's viewpoint.

* Corresponding author at: School of Chinese as a Second Language, Peking University, Yiheyuan Road 5, Beijing 100871, China.
E-mail address: chaosun@pku.edu.cn (C. Sun).

https://doi.org/10.1016/j.actpsy.2025.104908
Received 20 December 2024; Received in revised form 26 February 2025; Accepted 11 March 2025
Available online 14 March 2025
0001-6918/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Experimental research often investigates how perspective-taking influences the coordination of referential domains by manipulating perspective distinctions between participants. For example, certain information is shared by both participants, while other information is known only to one individual. In such cases, interlocutors often minimize or disregard their privileged knowledge, focusing instead on shared knowledge to facilitate communication (Barr, 2008; Barr & Keysar, 2002). These studies, by focusing on people's perspective-taking strategies during communication (Brown-Schmidt & Hanna, 2011; Hanna et al., 2003; Keysar et al., 1998, 2000), on the role of cognitive resources (Horton, 2005; Horton & Gerrig, 2016), and statistical models of perspective cues (Heller et al., 2016; Heller & Brown-Schmidt, 2023; Mozuraitis et al., 2018), have demonstrated that participants are able to adjust their expressions or understanding of dialogue based on shared knowledge between themselves and others.

1.2. Differentiation or integration of multiple perspectives

Previous research has mainly focused on dyadic communication, where information is categorized as either shared by two parties or privileged to one individual. However, when additional participants are involved, the processing of common ground becomes more complex due to the increased pairwise combinations of shared knowledge. This phenomenon raises the first question addressed in this study: when communicating with three or more partners, does an individual distinguish the perspective of each partner separately, or integrate the perspectives of all communicative partners?

The viewpoint supporting perspective integration argues that cognitively, differentiating the perspective of each individual partner is highly resource-demanding because language communication is rapid, and the amount of common ground information among three participants can be extensive. Therefore, individuals may also average the perspectives of all communicative partners in order to meet the communicative purpose with minimal cognitive resources.

This viewpoint is also supported by some experimental findings. Yoon and colleagues conducted a series of experiments to investigate perspective-taking strategies in multi-party conversations. In these studies, speakers shared different amounts of information with more than two addressees. Specifically, during a familiarization phase, speakers established certain conceptual pacts with some participants but not others. In the subsequent test phase, speakers were required to mention these concepts to all partners. The findings revealed that speakers adjusted their strategies depending on the ratio of knowledgeable partners. In particular, in three-party conversation, speakers often adopted the “aim-low” strategy, taking the perspective of the least knowledgeable participant (Yoon & Brown-Schmidt, 2014, 2018). However, as the number of participants increases, speakers tend to shift toward a “combined strategy,” where they take an average perspective of the group (Yoon & Brown-Schmidt, 2019). Specifically, participants exhibited more concise descriptions and increased fluency as the number of knowledgeable partners grew, showing greater consideration for the knowledgeable individuals. These studies highlight the variability of the speaker's audience design strategy depending on the number of interlocutors and the distribution of shared knowledge.

However, the aforementioned research evidence mainly comes from the speaker's audience design, that is, how speakers integrate the perspectives of all addressees to optimize their language expression. In contrast, addressees' discourse comprehension may involve perspective differentiation, because perspective-taking strategies can differ significantly between speakers and listeners. Speakers address all participants simultaneously in a conversation turn, whereas listeners process information from one speaker at a time. Consequently, addressees' perspective-taking strategies may exhibit less variation between dyadic and multi-party communication, as they may take the perspective of the specific individual who is speaking (Zheng & Sun, 2025).

Evidence of perspectives differentiation between two speakers has indeed been found in studies examining addressees' discourse comprehension in multi-party conversations, Yoon and Brown-Schmidt (2014) showed that addressees may differentiate between the perspectives of different partners because they interpret their partner's utterances with varying levels of fluency depending on the presence of uninformed participants. Zheng and Breheny (2021) conducted an experiment where an addressee faces two speakers who take turns giving instructions. In the study, speaker 1 always refer to the common ground between the addressee and her, while the referential domain of speaker 2 sometimes involves the addressee's privileged ground. The results showed that in subsequent communication, participants paid less attention to their privileged ground when interpreting speaker 1's utterances compared to interpreting speaker 2's utterances.

However, in the above studies, either communicators established common ground in independent dyadic communication, or the perspectives of the two communicative partners did not differ; only their referential domains in discourse differed. Therefore, the first experiment of the present study will establish a triadic communication context with inconsistent perspectives between two speakers based on a referential game, and investigate how the addressee interprets the two speakers' utterances. If, under consistent contextual conditions, participants interpret the same utterances from the two speakers differently, this would suggest the adoption of a differentiation strategy; if not, it would suggest an integration strategy.

1.3. The influence between multiple perspectives

However, even if addressees can understand the discourse according to the different perspectives of each speaker, it does not mean that these perspectives do not influence each other, nor does it prove that addressees do not perform partial integration. From a cognitive perspective, in multi-party communication, the more consistent the perspectives of the communicative partners, the less frequently individuals need to switch between perspectives. This allows them to have more cognitive resources to perform more accurate language comprehension and production. Previous cognitive processing theories on perspective-taking have emphasized the importance of episodic memory and priming. The memory-based model, for instance, highlights the role of episodic memory in tracking shared knowledge with specific conversational partners (Horton, 2005; Horton & Gerrig, 2016). According to this model, memories associated with a particular partner are more likely to be retrieved during subsequent interactions, but the co-presence of a conversational partner can also serve as a salient cue. When communicative partners are more homogeneous, they are more likely to serve as cues for memory retrieval. Similarly, priming mechanisms have been proposed to explain linguistic alignment in dialogue (Pickering & Garrod, 2004, 2013). When a speaker repeatedly uses a specific lexical item, their conversational partner is more likely to unconsciously adopt the same patterns due to memory priming. When the perspectives of communicative partners align, such priming occurs more frequently.

Furthermore, previous studies showing partner-specific processing have mainly focused on different perspectives between individuals and their communicative partners (Horton & Gerrig, 2002; Metzing & Brennan, 2003). However, in multi-party communication, the influence of different perspectives among communicative partners, or the degree of homogeneity among them, on individual processing remains underexplored. For example, in Horton and Gerrig's (2002) experiment, one instructor helped two participants sort cards by describing the contents of the cards. Each participant shared a different part of the card images with the instructor. The results showed that when the instructor described non-shared cards, they tended to adjust their speech in response when the participants changed, demonstrating partner-specific processing and supporting the idea of “participants distinguishing the perspectives of communicative partners.” However, studies like this have not answered whether differences in the perspectives of communicative partners affect participants' discourse production. For example, the same experiment also found that the audience design of participants differed depending on whether the two partners were present simultaneously. Only when the two partners were present separately did the audience design of the participants become prominent. This suggests that when communicative partners with different perspectives are co-present, their perspective processing might influence each other.

Therefore, whether perspectives between communicative partners influence each other will be the research motivation for the second experiment. In Experiment 2, the perspectives of the two speakers will be aligned. It is expected, first, that participants interpret the same discourse from the two speakers in the same way. Then, and crucially, this result will be compared with Experiment 1, where the discourse, physical context, and perspectives of speaker 2 remain unchanged, while the perspective of speaker 1 varies. If participants interpret the same discourse from speaker 2 differently, and the variation is similar to that of speaker 1, it would suggest that speaker 1's perspective-taking influenced participants' understanding of speaker 2's discourse.

1.4. The present study

The present study aims to investigate how addressees take multiple perspectives to interpret the utterances of speakers in three-party conversations. Specifically, the first experiment addresses the question of whether listeners distinguish between different perspectives and dynamically switch between them in alternating conversational turns, or whether they employ an integrated strategy, showing no distinct speaker-specific effects.

To explore this, the study employs a referential game in which interlocutors frequently refer to objects on a shelf. The visibility of the objects to each interlocutors are manipulated to examine how addressees use perspective cues to interpret language. Participants act as addressees in an object-picking task, where they are instructed to retrieve specific items from a shelf by two confederate speakers. The critical objects are labelled with Mandarin bare nouns, which lack determiners and allow both definite and indefinite interpretations (Li & Thompson, 1975, 1989). As a result, participants can either pick a specific object, corresponding to the definite interpretation, or any object, corresponding to the indefinite interpretation. The interpretation of the object will be influenced by the perspective constraints.

In Experiment 1, two speakers use the same verbal instructions in trials where the objects and their placement are identical. However, the two speakers hold different perspectives: Speaker 1 aligns with the participant addressee, seeing two potential candidates, while Speaker 2 sees only one. The participants' object choices and eye-tracking data related to the objects are monitored, as previous research has shown that participants' behavior reflects their explicit judgments, while eye-tracking data can reveal their implicit processing (Ferguson et al., 2017; Zheng & Sun, 2025). If the addressees select the same object under both speakers' instructions and show the same amount of gaze fixations toward both objects, it suggests that they have interpreted both speakers' utterances consistently and adopted an integrated perspective strategy. In contrast, if the addressees are more likely to choose and look at the object that only Speaker 2 can see under Speaker 2's instructions than under Speaker 1's, this suggests that they are adopting a perspective differentiation strategy.

The purpose of Experiment 2 is to test whether the perspectives of the two speakers interactively influence participants' processing. In Experiment 2, the two speakers have consistent perspectives, both of which align with Speaker 2 from Experiment 1, seeing only one candidate. First, it is expected that participants will interpret the two speakers' utterances similarly due to the same perspective cue. Next, the results of the two experiments will be compared, particularly regarding the interpretation of Speaker 2's utterances, as her perspective remains unchanged across the two experiments. If there is no change in participants' reference choices or gaze under Speaker 2's trial in both experiments, this indicates that the partner's perspective is independent and does not influence the other. However, if there are changes, and these changes follow a trend similar to Speaker 1's, it suggests that the perspective processing of the two partners can influence each other.

2. Experiment 1

2.1. Participants

Twenty-nine university students (age range: 18–23 years, average age 19.9 years; male: 13, female: 16) were recruited via campus advertisements to take part in the experiment. The sample size was determined based on previous studies with similar experimental designs in the field (Zheng & Breheny, 2021; Zheng & Sun, 2025). Additionally, two trained female experimental assistants served as confederate speakers. These assistants pretended to be participants who collaborated with the actual participants to complete the task. Participants were informed that the speaker role was assigned to them because they had acted as operators in a previous round.

All participants were right-handed and free from cognitive impairments, including hearing or visual difficulties. Prior to the experiment, they provided informed consent and were explicitly made aware of their right to withdraw at any point. The study was approved by the university's ethics review board, ensuring that all ethical guidelines were followed. All participants received monetary compensation, regardless of whether they completed the experiment.

2.2. Methods

2.2.1. Stimuli and design

During the experiment, participants performed an object-picking task on a shelf, following alternating instructions from two confederate speakers. The shelf consisted of a 2 × 2 grid, where three grids were transparent and one grid was blocked by a panel. Only individuals seated at the front of the shelf could see the contents of the blocked grid, while those seated at the back could not (Fig. 1). In this experiment, the participant and Speaker 1 sat in the front of the shelf, allowing both individuals to see all four objects. Speaker 2 sat at the back and could only see the objects in the three transparent grids.

At the start of each trial, the experimenter placed four objects into the four grids of the shelf. One speaker then instructed the participant to pick up an object from the shelf, using a sentence in the form of “请拿… (Please pick up...),” such as “请拿小狗 (Please pick up (the) dog).” The critical noun, such as “dog,” was presented as a bare noun, allowing the participant to interpret it either as a definite reference to a particular dog or as an indefinite reference to any dog.

The trials were divided into three types: **Non-Competitor Trials**: All four objects belonged to distinct categories (Fig. 1 left). **Competitor**

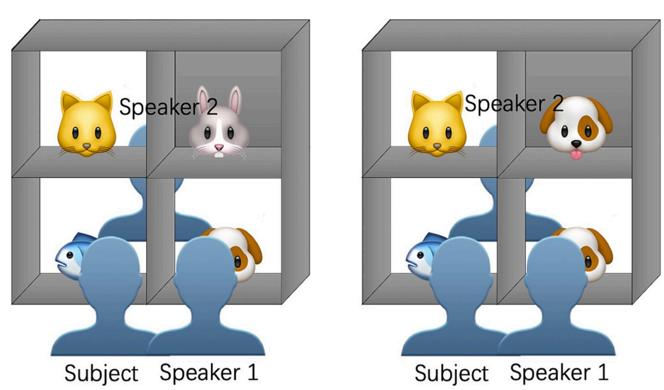


Fig. 1. The display of the shelf and the seating arrangement of participants and confederates in the non-competitor condition (left) and the competitor condition (right) in Experiment 1.

Trials: Three categories of objects were presented, with one category containing two identical items. One of these identical items was placed in the blocked grid (Fig. 1 right). **Filler Trials**: Similar to competitor trials, these trials included three categories of objects, one of which contained two identical items. However, in filler trials, the identical items were randomly distributed across the four grids and not restricted to the blocked grid. Furthermore, the speaker's instruction did not reference the identical objects. Filler trials were designed to prevent participants from developing strategies, such as anticipating that the speaker would mention identical objects whenever they appeared.

Each trial type included 48 critical trials, equally distributed between the two speakers (24 trials per speaker). For each speaker's 24 trials, there were 8 trials for each condition: 8 non-competitor trials, 8 competitor trials, and 8 filler trials. The objects and their placements in Speaker 1 and Speaker 2's trials were systematically matched, ensuring pairwise consistency. Similarly, non-competitor and competitor trials were matched by design, differing only in the presence or absence of identical objects. This resulted in eight trial combinations varying by speaker and competitor type. All trials were randomized into four sequences, and participants were assigned at random to one of these sequences.

In summary, the experiment followed a 2 × 2 within-subjects design, with Competitor Type (Non-Competitor vs. Competitor) and Speaker (Speaker 1 vs. Speaker 2) as factors. Each participant completed eight trials per condition.

2.2.2. *Procedure*

Before the experiment began, the experimenter explained the content and procedure of the experiment to the participant and the two confederate speakers. The experimenter described that the experiment consisted of a cooperative task and an individual hidden task. In the cooperative task, two instructors would give instructions to an operator to retrieve an object from a shelf. Each instructor had their own hidden task. The hidden task was explained to provide participants with an understanding of why and how the instructors gave orders.

Next, the experimenter assigned roles and clarified that since the two speakers had participated in the previous round of the game as operators, they would now take on the role of instructors in the current round. The experimenter then introduced the task requirements for each participant and speaker, particularly teaching the participants how to follow instructions and retrieve objects from the shelf.

Following this, the participants engaged in a practice session. In this phase, to better understand each other's perspectives, the participants took turns sitting on the back side of the shelf and on the front side of the shelf as a role of speaker. The two speakers also switched positions to experience the perspective differences from the other roles.

After completing the practice session, the formal experiment began. The participant and Speaker 1 sat on the front side of the shelf, while Speaker 2 sat on the back side. Depending on the assigned experimental order, the two speakers alternated giving instructions according to the order. To enhance the credibility of the experiment, the speakers pretended to be contemplating their hidden tasks while giving instructions. Additionally, the first five trials of the formal experiment were filler trials, designed to ensure that the participant was familiar with the procedure before engaging in the critical trials.

After finishing the experiment, the participant was asked whether they suspected that the speakers were experimental assistants or if they inferred the true purpose of the experiment. All participants reported that they did not make such inferences.

2.3. *Results*

2.3.1. *Behavior data analyses*

Participants' selections were categorized into three types: target, competitor, and other. Both the target and competitor objects represent the critical objects described by the noun; however, the target refers to the shared object, while the competitor refers to the blocked one. The "other" category includes instances where participants selected objects that were not mentioned or made no selection.

The experimenter recorded the objects chosen by participants during the test. Subsequently, a research assistant who was blind to the experimental design re-coded the participants' choices based on these recordings. The two sets of codings showed full consistency.

The proportions of participants' selection of the target and competitor objects are presented in Fig. 2. Speaker 1-competitor condition (M = 0.52, SD = 0.50) was numerically lower and closer to the probabilistic level (0.5) than the Speaker 2-competitor condition (M = 0.66, SD = 0.48). There were no non-target selections recorded for either Speaker 1 or Speaker 2 in the non-competitor condition. Participants also rarely selected objects outside the target and competitor categories, indicating target and competitor selections were complementary. Accordingly, only analyses of target selections were performed, and further analyses of competitor selection were deemed unnecessary.

A Bayesian generalized linear mixed-effects model (GLMM) was conducted using R with the *brms* function. The model was fitted to examine the effects of Competitor Types (competitor vs. non-competitor), Speaker (Speaker 1 vs. Speaker 2) and their interaction on the target selections, with a maximal random effects structure (random intercepts and slopes) for both subjects and items. The model used a Bernoulli distribution for the binary outcome and Bayesian MCMC sampling for parameter estimation. The model had good convergence, with Rhat values of 1.00 for all parameters, indicating that the chains successfully converged.

The analysis revealed a significant effect of Competitor Type, with target selections being higher in the non-competitor condition compared to the competitor condition, $\beta = 9.10$, SE = 2.35, 95 % CI [5.46, 14.45]. There was also a significant main effect of Speaker, with higher target selections observed in Speaker 2 compared to Speaker 1, $\beta = 0.92$, SE = 0.43, 95 % CI [0.11, 1.81], showing that participants distinguished the perspectives of both speakers However, the interaction between Competitor Type and Speaker2 was not statistically significant, $\beta = 2.71$, SE = 3.39, 95 % CI [−3.59, 9.89].

2.3.2. *Eye-tracking data analyses*

Similar to the behavioral data, participants' fixations were categorized into target, competitor, and other. When participants fixated on the grid containing the target object, their fixation was defined as "target". Similarly, "competitor" refers to when participants fixated on the blocked grid, while "other" refers to all other cases except these two. Eye movement data were recorded at 25 frames per second, meaning the data were updated every 40 ms time bin. Participants' fixations were analyzed between 200 and 1200 milliseconds after hearing the critical word. The 200 ms starting point was chosen based on the average time for launching eye movement (Hallett, 1986), while the 1200 ms

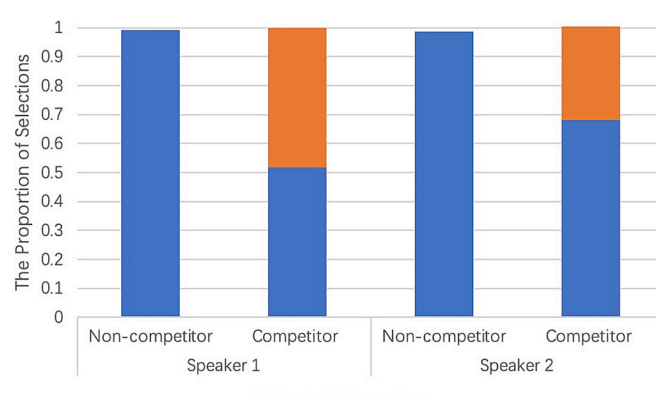


Fig. 2. The proportion of participants' selection of the target and competitor objects in each condition in Experiment 1.

endpoint was chosen based on the average time it took participants to stop fixating on the shelf and begin to pick up the objects. Fig. 3 shows participants' fixation trajectories toward the target and competitor objects across two competitor types and two speaker conditions. The analyses were conducted on target and competitor fixations separately.

2.3.2.1. Target analysis. A general linear mixed model was conducted using R with the *lmer* function. Competitor Types (competitor vs. non-competitor) and Speaker (Speaker 1 vs. Speaker 2) were entered into the model as fixed effects, with a maximal random effects structure (random intercepts and slopes) for both subjects and items. The target fixations served as the dependent variable.

There was a significant main effect of Competitor Type, $\beta = 0.356$, SE $= 0.112$, $t = 3.168$, $p = .003$, suggesting that participants had more fixations in the non-competitor condition compared to the competitor condition. There was no significant Speaker effect, $\beta = 0.065$, SE $= 0.058$, $t = 1.132$, $p = .266$, and no interaction between Competitor Type and Speaker, $\beta = -0.078$, SE $= 0.072$, $t = -1.078$, $p = .290$. The result indicates that participants might take the perspectives of both speakers as a whole. They did not fixate more on the shared target in Speaker 2's trials simply due to her perspective.

2.3.2.2. Competitor analysis. A general linear mixed model was performed with Competitor Types (competitor vs. non-competitor) and Speaker (Speaker 1 vs. Speaker 2) as fixed effects, with random intercepts and slopes for both subjects and items. The competitor fixations served as the dependent variable.

There was a significant main effect of Competitor Type, $\beta = -0.379$, SE $= 0.078$, $t = -4.875$, $p = .000$, indicating that fixations on competitor objects were significantly higher in the competitor condition compared to the non-competitor condition. There was no significant effect of Speaker, $\beta = -0.099$, SE $= 0.051$, $t = -1.945$, $p = .060$, but there was a trend toward greater competitor fixations for Speaker 1 trials. No interaction between competitor type and speaker was found, $\beta = 0.077$, SE $= 0.048$, $t = 1.608$, $p = .119$. This result supports a slightly distinct processing of Speaker 1 and 2's trials, given the result that participants had a tendency to fixate more on the competitor in Speaker 1 trials.

2.3.2.3. Comparison between target and competitor. The Wilcoxon signed-rank test was performed to compare the competitor and target fixations in the competitor condition for Speaker 1 and Speaker 2 separately in R with the *wilcox.test* function. There was a significant difference between competitor and target in both Speaker 1, V $=$ 3,157,946, $p < .001$, and Speaker 2, V $=$ 2,216,412, p $=$ .000, indicating significantly higher fixation on the target in both Speaker trials. Participants fixated more on the target, even though in Speaker 1 trials where both target and competitor were visible to Speaker 1, which supports an integrated strategy and suggests that the processing of Speaker 1's perspective is influenced by Speaker 2 to some extent.

2.4. Discussion

In the experiment, the participants collaborated with two speakers on an object-picking task on grids, with Speaker 1 sharing the same perspective with the participant, while Speaker 2 was unaware of the information in the blocked grid. Behavioral data showed that participants interpreted the bare nouns of each speaker's instruction based on each speaker's perspective, respectively. For Speaker 1, whose perspective involved two potential referential candidates for the bare nouns, participants made indefinite interpretations, choosing both the shared and unshared objects near a probabilistic level (0.5). However, for Speaker 2, participants adopted her perspective to narrow down the referential domain of the nouns, thus making more definite interpretations of the bare nouns and choosing the shared object more often than the unshared objects. Overall, the target preference was significantly higher for Speaker 1 than for Speaker 2.

Eye-tracking data showed that participants attended to the shared target objects equally across both speaker trials. This similarity may have resulted from the fact that all three interlocutors shared the target objects. In contrast, participants experienced greater interference from the competitor objects in Speaker 1's trials compared to Speaker 2's trials, due to the fact that they shared the competitors with Speaker 1 rather than Speaker 2. Notably, they attended more to the target than the competitor in Speaker 1's trials, showing the integration of Speaker 2's perspective.

In sum, the study indicates that participants explicitly distinguish the perspectives of different speakers during the referential resolution of Mandarin bare nouns, but the eye-movement data suggest an implicit integration of perspectives. Moreover, the second question raised at the beginning-how the perspectives between the two speakers interactively influence participants' interpretation-remains unclear. In the following experiment, two speakers will share the same grids but be ignorant of the blocked grid, and the difference between the two experiments will be compared to explore the role of partner similarity.

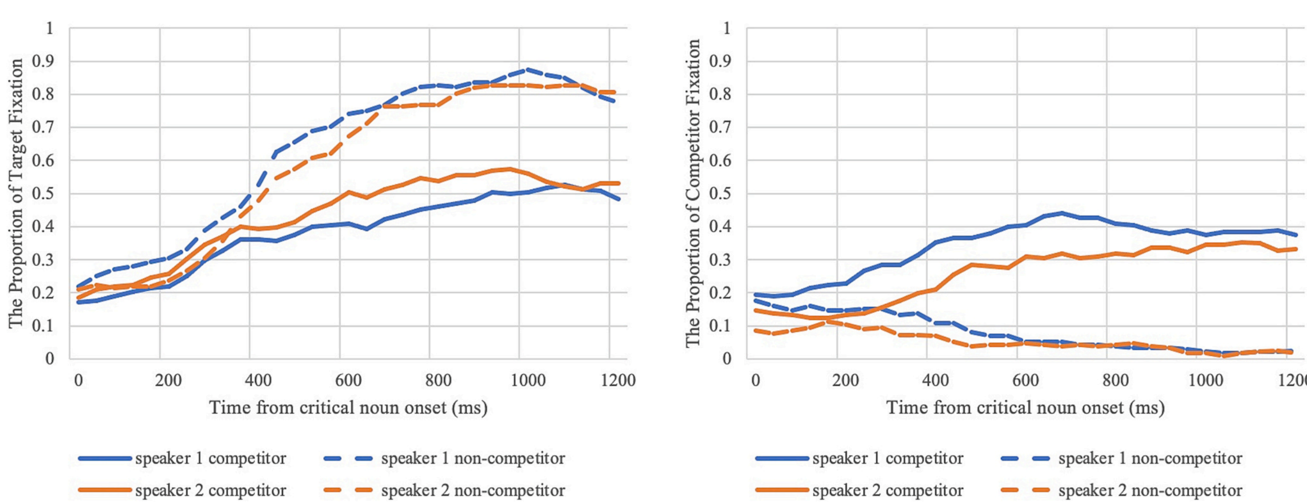


Fig. 3. The 1200-ms time window capturing participants' eye fixations to target objects (upper) and competitor objects (lower) from the onset of the critical noun in Experiment 1.

3. Experiment 2

3.1. Participants

Twenty-one university students (age range: 18–24 years, average age 19.6 years; male: 10, female: 11) participated in Experiment 2. The recruitment method and participant requirements for participants were the same as in Experiment 1, and the rewards for participation was identical. The two assistants from Experiment 1 continued to serve as confederate speakers.

3.2. Methods

In the previous experiment, Speaker 1 not only shared the information with the participants, but also physically close to them. To maintain the consistency of the effect of physical proximity, the positions of the speakers in the second experiment were the same as in the first, with only the transparency of the grids being altered. In Experiment 2, the blocked grid had panels on both the front and back sides, but the panel on the front side did not fully cover the grid, allowing the participants to see the objects within the grid from their perspective, while Speaker 1 could not see them.

The procedure for Experiment 2 was identical to that of Experiment 1. After arriving at the lab, participants completed practice sessions from different positions and perspectives. Subsequently, the formal test began. Experiment 2 used the same objects and their arrangement as in Experiment 1, and the trial order followed the same sequence as in Experiment 1 (Fig. 4).

3.3. Results

3.3.1. Behavior data analyses

The behavioral data were coded in the same manner as Experiment 1. The consistency between the two coders was 100 %. The proportion of participants' selection of the target and competitor objects is presented in Fig. 5. Speaker 1-competitor condition (M = 0.81, SD = 0.39) and Speaker 2-competitor condition (M = 0.86, SD = 0.35) are both numerically greater than the probabilistic level (0.5). As in Experiment 1, there were no non-target selections recorded in the non-competitor condition, and target and competitor selections were complementary in the competitor condition.

A Bayesian generalized linear mixed-effects model (GLMM) examined the effects of Competitor Type, Speaker, and their interaction on participants' target selections, with the maximal random effects structure. There was a significant effect of Competitor Type, with target selections being higher in the non-competitor condition compared to the competitor condition, β = 7.69, SE = 2.47, 95 % CI [3.57, 13.40]. However, the main effect of Speaker was not statistically significant, β = 0.49, SE = 0.45, 95 % CI [−0.39, 1.38]. The interaction between Competitor Type and Speaker also did not show a significant effect, β = 2.42, SE = 3.49, 95 % CI [−3.70, 9.77]. The results suggest that participants based their interpretation on the speaker's visual perspective rather than their physical position.

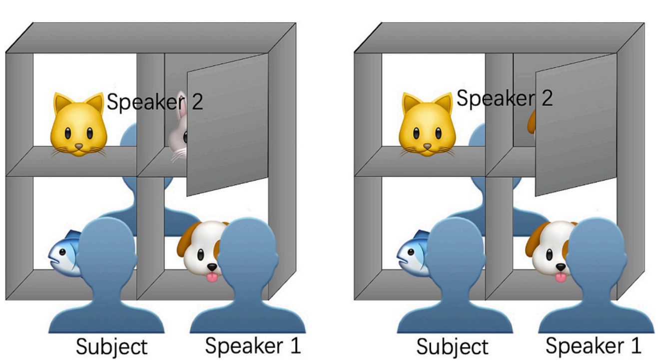


Fig. 4. The display of the shelf and the seating arrangement of participants and confederates in the non-competitor condition (left) and the competitor condition (right) in Experiment 2.

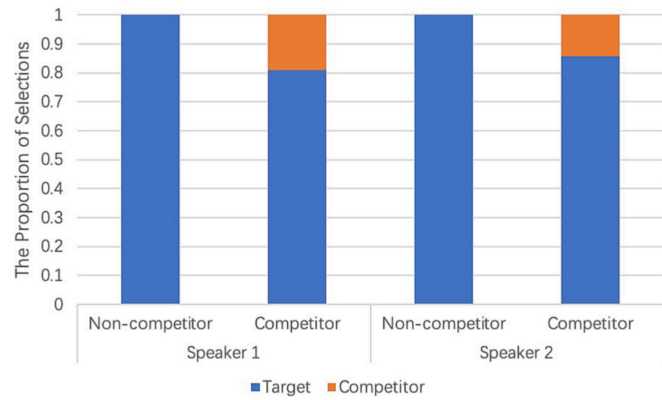


Fig. 5. The proportion of participants' selection of the target and competitor objects in each condition in Experiment 2.

3.3.2. Eye-tracking data analyses

3.3.2.1. Target analyses. The eye-tracking data were coded and analyzed using the same methods as in Experiment 1. A general linear mixed-effects model was fitted to examine the effect of Competitor Type, Speaker, and their interaction on the target fixation for each 40 ms bin, with the maximal random effects structure.

There was no main effect of Competitor Type, β = 0.076, SE = 0.093, t = 0.814, p = .424, no main effect of Speaker, β = 0.012, SE = 0.046, t = 0.272, p = .788, and no interaction effect, β = 0.007, SE = 0.060, t = 0.130, p = .898 (Fig. 6).

3.3.2.2. Competitor analyses. A general linear mixed-effects model was also performed to examine the effect of Competitor Type, Speaker, and their interaction on the competitor fixation for each 40 ms bin, with the maximal random effects structure. The results showed that there was a significant main effect of Competitor types, β = −0.169, SE = 0.076, t = −2.204, p = .038, due to greater competitor fixation in the competitor condition than in the non-competitor condition. There was no significant Speaker effect, β = −0.055, SE = 0.043, t = −1.290, p = .208, and no interaction between competitor types and speaker, β = 0.020, SE = 0.038, t = 0.526, p = .604 (Fig. 6).

3.3.3. Combined analysis of Experiments 1 and 2

3.3.3.1. Behavioral data analyses. The following analyses focused on whether participants displayed different performances across two experiments in Speaker 1 and Speaker 2 trials, respectively. Bayesian generalized linear mixed-effects models (GLMM) were used to examine the effects of Competitor Type (competitor vs. non-competitor) and Experiment (1 vs. 2) on participants' target selections with the maximal random effects structure.

For the Speaker 1 trials, there was a significant effect of Competitor Type, with target selections being higher in the non-competitor condition compared to the competitor condition, β = 8.87, SE = 2.35, 95 % CI [5.26, 14.06]. Additionally, there was a significant main effect of Experiment, with higher selections observed in Experiment 2 compared to Experiment 1, β = 2.01, SE = 0.60, 95 % CI [0.95, 3.31]. The interaction between Competitor Type and Experiment was not statistically significant, β = 2.17, SE = 3.49, 95 % CI [−4.15, 9.45]. This finding is consistent with the prediction that Speaker 1 could not see the

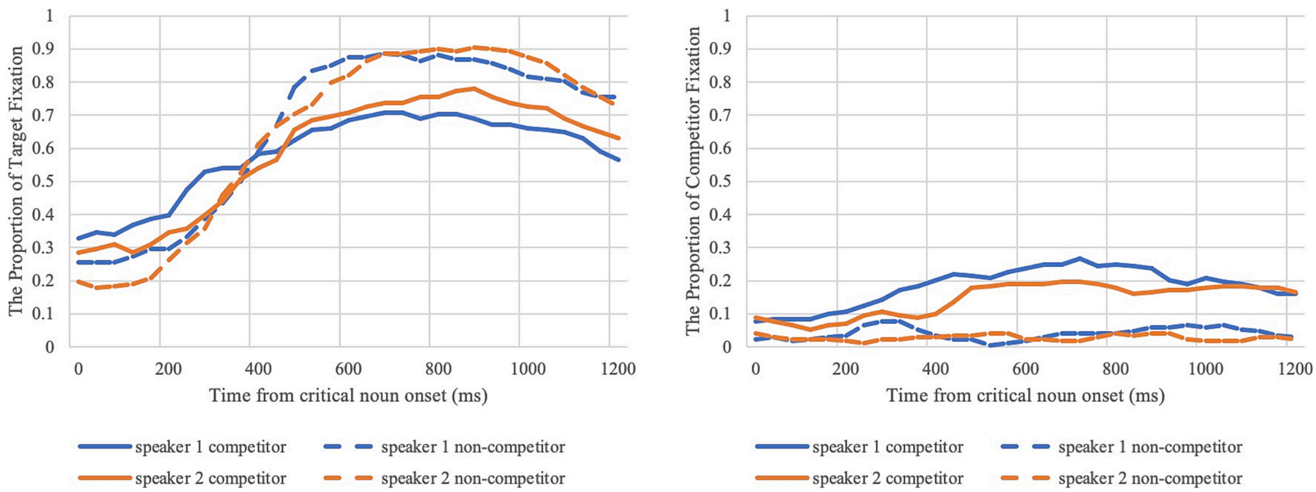


Fig. 6. The 1200-ms time window capturing participants' eye fixations to target objects (upper) and competitor objects (lower) from the onset of the critical noun in Experiment 2.

competitor object in the second experiment, and participants took her perspective.

A similar pattern was acquired in Speaker 2 trials. A significant effect of Competitor Type was found, showing higher target selections in the non-competitor condition relative to the competitor condition, $\beta = 8.56$, SE = 2.28, 95 % CI [4.85, 13.72]. There was also a significant main effect of Experiment, with higher target selections observed in Experiment 2 compared to Experiment 1, $\beta = 1.51$, SE = 0.67, 95 % CI [0.27, 2.92]. The interaction between Competitor Type and Experiment was not statistically meaningful, $\beta = 2.05$, SE = 3.52, 95 % CI [−4.39, 9.56]. The result suggests that the perspectives of both speakers interactively influence participants' interpretation, because the interpretation of Speaker 2 followed the same pattern as that of Speaker 1.

3.3.3.2. Eye-tracking data analyses. The following analyses focused on the target and competitor fixations across two experiments in Speaker 1 and Speaker 2 trials separately. General linear mixed-effects models were used to examine the effects of Competitor Type (competitor vs. non-competitor) and Experiment (1 vs. 2) on the target and competitor fixations, using the maximal random effects structure.

For Speaker 1 target analyses, the results showed that there was a significant main effect of Competitor Type, $\beta = 0.472$, SE = 0.109, $t = 4.339$, $p = .000$, with greater target fixation in the non-competitor condition compared to the competitor condition. There was a significant effect of Experiment, $\beta = 0.223$, SE = 0.072, $t = 3.096$, $p = .006$, indicating that participants had a larger proportion of target fixation in Experiment 2 compared to Experiment 1. There was also an interaction between Competitor Type and Experiment, $\beta = -0.193$, SE = 0.067, $t = -2.879$, $p = .008$. Post-hoc analysis revealed that in the non-competitor condition, there was no difference between Experiment 1 and Experiment 2, $\beta = -0.030$, SE = 0.050, $z = -0.593$, $p = .553$. However, in the competitor condition, there was a greater target preference in Experiment 2, $\beta = -0.223$, SE = 0.072, $z = -3.096$, $p = .002$.

For Speaker 1 competitor analyses, there was a significant main effect of Competitor Type, $\beta = -0.455$, SE = 0.087, $t = -5.254$, p = .000, indicating greater competitor fixation in the competitor condition compared to the non-competitor condition. A significant effect of Experiment was also found, $\beta = -0.161$, SE = 0.061, $t = -2.651$, $p = .014$, suggesting that participants had a lower proportion of competitor fixation in Experiment 2 compared to Experiment 1. Furthermore, an interaction between Competitor type and Experiment was found, $\beta = 0.153$, SE = 0.055, $t = 2.796$, $p = .010$. Post-hoc analysis indicated that no significant difference was found in the non-competitor condition between the two experiments, $\beta = 0.008$, SE = 0.025, z = 0.322, p = .008, $p = .748$. However, in the competitor condition, there was a significant difference between Experiment 1 and Experiment 2, $\beta = 0.161$, SE = 0.061, z = 2.651, p = .008, demonstrating less competitor interference in Experiment 2.

For Speaker 2 target analyses, there was a significant Competitor Type effect, $\beta = 0.312$, SE = 0.127, $t = 2.454$, $p = .021$, indicating greater target fixation in the non-competitor condition compared to the competitor condition. There was also a significant effect of Experiment, $\beta = 0.177$, SE = 0.076, $t = 2.317$, $p = .029$, indicating that participants showed a larger proportion of target fixation in Experiment 2. However, no significant interaction was observed between Competitor Type and Experiment, $\beta = -0.111$, SE = 0.082, $t = -1.361$, $p = .189$.

For Speaker 2 competitor analyses, there was a significant effect of Competitor Type, $\beta = -0.321$, SE = 0.090, $t = -3.551$, $p = .001$, indicating that competitor fixations were greater in the competitor condition compared to the non-competitor condition. There was no significant effect of Experiment, $\beta = -0.115$, SE = 0.058, $t = -1.987$, $p = .055$, but a trend suggested that participants tended to show fewer competitor fixations in Experiment 2. There was no significant interaction between Competitor Type and Experiment, $\beta = 0.095$, SE = 0.052, $t = 1.840$, $p = .076$.

In summary, the participants exhibited similar patterns of referential interpretation for the discourse of Speaker 1 and Speaker 2 in Experiment 2: compared to Experiment 1, participants demonstrated a greater preference for the target and less interference from competitors in Experiment 2.

3.4. Discussion

In Experiment 2, although Speaker 1 and Speaker 2 were seated in different positions, they could see the same objects. Behavioral data showed that participants interpreted the utterances of both speakers in the same way, and eye-tracking data also indicated that participants exhibited similar attention patterns to referential candidates during utterance comprehension. This suggests that participants relied on shared knowledge with the speakers rather than on physical distance when interpreting bare nouns.

Next, a comparison between Experiments 1 and 2 reveals that the changes in how participants interpreted Speaker 1's utterances corresponded with changes in her field of view. Specifically, in Experiment 2, participants were more likely to focus on the shared target object and showed reduced interference from competitors, because Speaker 1 could not see the competitor objects in the blocked grid. This cross-experiment comparison further supports the conclusion that participants relied on perspective cues to narrow the referential domain.

Interestingly, such changes not only occurred in their responses to Speaker 1 but also extended to Speaker 2. The pattern of change in interpreting Speaker 2's utterances across the two experiments was similar to that of Speaker 1: participants experienced less interference from competitors and demonstrated larger target preference in Experiment 2. Notably, Speaker 2's position and visual field remained unchanged across the two experiments, and the objects and order of instructions were identical, ensuring comparability. These results suggest that participants' interpretation of Speaker 1's utterances influenced their interpretation of Speaker 2's utterances. This influence was not only evident in their explicit behavioral choices but also reflected in distinct patterns of eye gaze toward shared and unshared objects.

In conclusion, across Experiments 1 and 2, when participants interacted with two speakers simultaneously, they were able to differentiate between perspective information. Meanwhile, participants' adoption of one speaker's perspective influenced their adoption of the other speaker's perspective, ultimately shaping specific referential resolution strategies during different conversations.

4. General discussion

This study investigates how the addressee utilizes the perspectives of speakers to resolve referential ambiguity in three-party communication. Specifically, it examines whether addressees differentiate between the perspectives of two speakers and whether the perspectives of the two speakers influence each other. In two experiments, participants, acting as addressees, collaborated with two interlocutors on an object-picking task. The task involved resolving referential expressions in the speakers' instructions to identify the target object. Participants' referential choices provided insight into their interpretation of their partners' referential domains and how these domains were constrained by the speakers' perspectives. The referential expressions employed Mandarin bare nouns, which can denote a specific object (definite interpretation) or any object that meets the semantic concept (generic interpretation). In both experiments, participants always viewed two identical referential candidates, which were assumed to have a generic interpretation, while Speaker 2 had access to a single object, which supported a definite interpretation. In Experiment 1, Speaker 1 shared the same perspective as the participant, thus supporting the generic interpretation, whereas in Experiment 2, Speaker 1 shared the same perspective as Speaker 2, thus supporting the definite interpretation.

4.1. *Independent processing of perspectives*

Behavioral results from Experiment 1 demonstrated that participants independently took the perspective of each speaker during referential resolution, with a definite interpretation for Speaker 2 but a generic interpretation for Speaker 1's utterance. However, in Experiment 2, where both speakers shared identical perspectives, participants exhibited similar behavioral and eye fixation patterns when processing instructions from the two speakers, showing a predominance of definite interpretations for both speakers.

This partner-specific strategy differed from the audience design in multiparty communication, where speakers adopt either the least knowledgeable perspective or average the perspectives across all interlocutors to enhance comprehension efficiency (Yoon & Brown-Schmidt, 2014, 2018, 2019). The reason for this difference lies in the number of partner(s) a speaker or an addressee faces. In language production, a speaker faces several addressees with different perspectives, whereas in comprehension, an addressee faces a specific individual with a particular perspective. Accordingly, participants' behavior aligns more closely with partner-specific processing observed in dyadic communication, where individuals consider distinct common grounds for different partners (Brennan & Hanna, 2009; Brown-Schmidt, 2009; Metzing & Brennan, 2003). In such studies, participants established conceptual pacts with a specific partner, and when that partner used new expressions, their comprehension exhibited delays. However, no such delays occurred when a new conversational partner introduced new expressions (Brown-Schmidt, 2009; Metzing & Brennan, 2003). Similarly, participants also established partner-specific referential domains with different speakers. For example, in Zheng and Breheny's experiments, two speakers and the addressee had different referential domains based on prior discourse, and participants gradually established referential domains specific to each speaker (Zheng & Breheny, 2021).

4.2. *Integrated processing of perspectives*

However, a comparison across experiments reveals that, despite Speaker 2's perspective remaining constant, participants in Experiment 2 showed a stronger preference for the shared target object and were less influenced by their privileged competitor when Speaker 2 gave instructions, compared to Experiment 1. This suggests improved perspective-taking for Speaker 2 in Experiment 2. This finding can be interpreted as either indicating that participants partially integrated the perspectives of Speaker 1 and Speaker 2 when distinguishing between them, or that participants' perspective-taking for Speaker 1 influenced their perspective-taking of Speaker 2. The result can be explained by the theoretical framework as follow:

4.2.1. *The interpretation of the speaker's strategy*

The results imply that addressee participants may have accounted for the speakers' audience design strategies. When speakers take the perspective of either the least knowledgeable partner or an averaged perspective of all partners in multi-party communication, addressees may have adjusted their understanding accordingly. That is, they may have used the same strategy to interpret. In dyadic conversation, the common ground is established based on the awareness of the shared knowledge, meaning that each party knows that the other knows the knowledge. Both interlocutors dynamically establish the information as common ground as it becomes overt to both parties (Aumann, 1976; Sperber & Wilson, 1986). It can be assumed that in multi-party conversation, the awareness of the shared knowledge is replaced by the strategy used in the application of the shared knowledge. Participants may evaluate how each party treats the shared knowledge, and dynamically converge toward one particular coordinated strategy, with a unified criterion on multiple perspectives.

Eye-tracking data from Experiment 1 also support this interpretation: participants showed no differences in eye fixation on the shared target object when interpreting the utterances of the two speakers. This result may suggest that the addressee's attentional domain tends to align when interpreting the utterances of both speakers. Furthermore, during Speaker 1 trials, participants displayed greater fixation on the shared target object compared to the competitor, indicating some consideration of Speaker 2's ignorance regarding the competitor.

4.2.2. *Grounding by proxy*

Another explanation for the results is that participants tended to view certain communicative partners as a collective unit and process their perspectives as a whole. Previous research has shown that, in multi-party conversations, as the topic progresses, people gradually form a team through verbal and non-verbal cues (Kangasharju, 1996). This team formation in multi-party communication may alter individual-level perspective-taking. Eshghi and Healey (2016) proposed the "Grounding by Proxy" mechanism in multi-party dialogue, where one person's response can serve as a proxy for another's within a conversational team or alliance. Furthermore, this team formation may also be pre-existing. For example, people often assume that individuals from the same culture or family share certain information, which is referred to as cultural common ground (Clark, 1996). Even for two strangers who have never met before, cultural common ground can be used to infer the meaning of their discourse. In other cases, such as in a task-oriented setup, the roles of the task performers may lead them to perceive the task director as an alliance. Specifically, in the present study, the two speakers have similar tasks, both instructing the object-picking process, making it easier for the two to be interpreted as a team.

During Experiment 1, although Speaker 2 lacked certain object information, Speaker 1 could proxy for both speakers, leading to participants' overestimating Speaker 2's knowledge. In contrast, participants in Experiment 2 demonstrated better perspective-taking of Speaker 2 and were less influenced by their privileged information. This can be explained by the two speakers sharing a more consistent perspective, reinforcing their alliance.

However, this explanation may have limited applicability due to the specific nature of the task types. In task-based dialogues, alliance-based relationships may be more effective, but in conversations regarding information or personal emotions, individuals may prioritize individual differences and adopt different strategies (Dideriksen et al., 2023).

4.2.3. Probabilistic calculations of multiple perspectives

A third explanation is that this strategy may be based on probabilistic calculations of multiple factors, ultimately leading individuals to adopt a more economical and effective perspective-taking strategy that benefits a larger number of people. Research has suggested that in dyadic conversations, individuals process perspectives by assigning different weights to multiple perspectives (e.g., self and other), based on the context and linguistic expressions. This involves Bayesian computations of perspective information (Heller et al., 2016; Heller & Brown-Schmidt, 2023). Such phenomena can also occur in multi-party communication. In multi-party settings, people form representations of each participant's perspective, and the final perspective-taking strategy may be influenced by the number of participants and the degree of consistency among the perspectives.

In Experiment 2, there was greater consistency between the two speakers, whereas in Experiment 1, Speaker 1's perspective was more aligned with that of the participant. The results implied that in Experiment 1, the stronger alignment between the participant and Speaker 1 might have caused participants' privileged ground to be weighted more heavily, which subsequently extended to interpretations of Speaker 2's utterances. Hence, participants relied more on their egocentric perspective in Experiment 1 than in Experiment 2 when interpreting Speaker 2's discourse.

This result also aligns with previous research which has shown that while shared knowledge between interlocutors facilitates communication, high overlap can hinder it due to overestimation of common ground (Wu & Keysar, 2007). People tend to exhibit egocentric biases more strongly with familiar partners than with strangers, often overestimating their understanding of familiar individuals' mental states (Robbins & Krueger, 2005; Savitsky et al., 2011). It is possible that the alignment between acquaintance and oneself increases the weight of privileged ground information.

4.3. The cognitive resource

The perspective-taking strategy may also depend on cognitive resources. When people have sufficient cognitive resources to process each individual's perspective, they are more likely to differentiate and understand each person's discourse. As the number of participants increases, however, and with limited time to process all perspectives, people are more likely to adopt an integrated strategy. Meanwhile, higher consistency among participants may reduce cognitive demands, while greater differentiation increases processing complexity.

In this study, participants interacted with only two communicative partners, so cognitive resources were adequate, resulting in perspective differentiation. Nonetheless, the findings still revealed integration in participants' processing of the two speakers' perspectives. In Experiment 1, participants' perspective-taking for Speaker 1 led to greater retrieval of competitor-related memory traces during Speaker 2's utterances compared to Experiment 2. This influence was particularly evident in implicit eye-tracking data. In Experiment 1, Speaker 1's behavioral choices were probabilistic, but participants' eye fixation on the target objects visible to all three individuals was greater than their fixation on objects in the blocked grid, suggesting mutual influence between the two speakers' perspectives. This interaction may occur through a low-level, automatic process: when two speakers alternated in addressing the participant, the linguistic and visual-related representations associated with one speaker could be triggered by the other co-present speaker.

Horton and Gerrig (2005, 2016) proposed a domain-general, memory-based model that suggests events are stored in episodic memory, and during subsequent communication, cues associated with these memories, such as the individuals present at the time of the event, trigger memory retrieval. Similarly, Pickering and Garrod (2004, 2013, 2021) noted that alignment in dialogue, such as the use of similar lexical, syntactic, or phonological forms, occurs through a specific processing mechanism called “memory priming.” While the present study focused primarily on the real-time integration of visual and linguistic information, similar attentional reinforcement or priming effects were also observed at the attentional level.

Such priming is automatic rather than strategic and does not involve higher-level social cognition or control mechanisms. Unlike the aforementioned social cognitive strategies, which are more likely to be influenced by the number of participants and the similarity of perspectives, low-level automatic processing may be affected by the salience of retrieval cues or the strength of their associations. The design of future studies can address these questions.

5. Conclusion

This study explored whether and how participants, as addressees, consider the perspectives of two speakers while resolving referential ambiguity in a collaborative task. The results revealed that participants distinguished between speakers' perspectives and provided different interpretations of the same noun accordingly. However, participants also integrated perspectives to some extent. This potential strategy was discussed in terms of considerations of their partner's audience design strategies, using a “Grounding by Proxy” strategy, or a probabilistic calculation of different perspectives. Low-level memory priming was also discussed based on the implicit results from the eye-tracking data. However, these strategies may be influenced by communicative goals and group size. Further research is needed to examine perspective-taking in different communicative contexts, with varying group sizes and degrees of perspective overlap.

CRediT authorship contribution statement

Xiaobei Zheng: Writing – original draft, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Chao Sun:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Data curation, Conceptualization.

Funding

This work was supported by the Peking University Humanities and Social Sciences Starting Fund.

Declaration of competing interest

The authors declare none.

Appendix A

Appendix 1
Objects used in each trial of competitor condition.

Trail no.	other	other2	target	competitor
1	apple	car	dog 1	dog 2
2	airplane	fish	moon 1	moon 2
3	cat	flower	watermelon 1	watermelon 2
4	tree	grape	train 1	train 2
5	apple	airplane	cat 1	cat 2
6	ship	rabbit	ball 1	ball 2
7	moon	dog	banana 1	banana 2
8	flower	grape	car 1	car 2

Appendix 2
Objects used in each trial of competitor condition.

Trail no.	other	other2	target	competitor
1	apple	car	dog	ball
2	airplane	fish	moon	banana
3	cat	flower	watermelon	ship
4	tree	grape	train	rabbit
5	apple	airplane	cat	tree
6	ship	rabbit	ball	banana
7	moon	dog	banana	train
8	flower	grape	car	fish

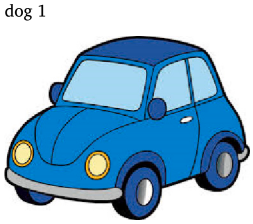
Appendix 3
The images used in competitor and non-competitor condition.


dog 1


dog 2

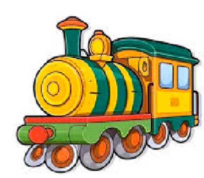

cat 1


cat 2


car 1

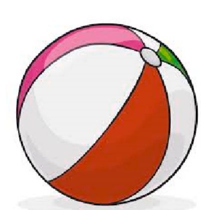

car 2

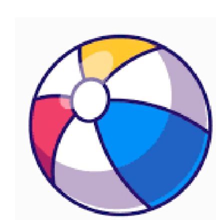

train 1


train 2


moon 1


moon 2

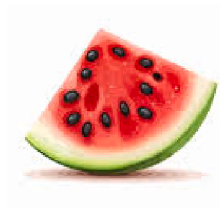

ball 1


ball 2

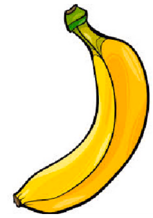
(*continued on next page*)

Appendix 3 (*continued*)

watermelon 1



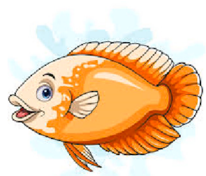
watermelon 2



banana 1

banana 2

rabbit



fish

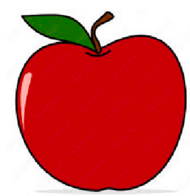
ship

plane

tree

grape

flower



apple

Data availability

Data will be made available on request.

References

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics, 4*(6), 1236–1239.
Barr, D., & Keysar, B. (2002). Anchoring and comprehension in linguistic precedents. *Journal of Memory and Language, 46*, 391–418.
Barr, D. J. (2008). Pragmatic expectations and linguistic evidence: Listeners anticipate but do not integrate common ground. *Cognition, 109*(1), 18–40.
Beun, R. J., & Cremers, A. H. (1998). Object reference in a shared domain of conversation. Pragmatics & Cognition, 6(1–2), 121–152.
Brennan, S. E., & Hanna, J. E. (2009). Partner-specific adaptation in dialog. *Topics in Cognitive Science, 1*(2), 274–291.
Brown-Schmidt, S. (2009). Partner-specific interpretation of maintained referential precedents during interactive dialog. *Journal of Memory and Language, 61*(2), 171–190.
Brown-Schmidt, S., & Hanna, J. E. (2011). Talking in another person's shoes: Incremental perspective-taking in language processing. *Dialogue & Discourse, 2*(1), 11–33.
Clark, H. H. (1996). *Using language*. Cambridge: Cambridge University Press.
Clark, H. H., & Brennan, S. E. (1991). Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley (Eds.), *Perspectives on socially shared cognition* (pp. 127–149). American Psychological Association.
Clark, H. H., & Clark, E. V. (1977). *Psychology and language: An introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
Clark, H. H., & Marshall, C. R. (1978). Reference diaries. In D. L. Waltz (Ed.), *Vol. 2. Theoretical issues in natural language processing* (pp. 57–63). New York: Association for Computing Machinery.
Clark, H. H., & Marshall, C. R. (1981). Definite reference and mutual knowledge. In A. K. Joshi, B. L. Webber, & I. A. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge University Press.
Clark, H. H., & Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition, 22*(1), 1–39.
Dideriksen, C., Christiansen, M. H., Tylén, K., Dingemanse, M., & Fusaroli, R. (2023). Quantifying the interplay of conversational devices in building mutual understanding. *Journal of Experimental Psychology: General, 152*(3), 864.
Eshghi, A., & Healey, P. G. (2016). Collective contexts in conversation: Grounding by proxy. *Cognitive Science, 40*(2), 299–324.
Ferguson, H. J., Apperly, I., & Cane, J. E. (2017). Eye tracking reveals the cost of switching between self and other perspectives in a visual perspective-taking task. *Quarterly Journal of Experimental Psychology, 70*(8), 1646–1660.
Gerrig, R. J., & Horton, W. S. (2004). Contextual expressions and common ground. In *Figurative language comprehension* (pp. 43–70). Routledge.
Gildea, P., & Glucksberg, S. (1983). On understanding metaphor: The role of context. *Journal of Verbal Learning and Verbal Behavior, 22*(5), 577–590.
Hallett, P. E. (1986). Eye movements. In K. B. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance I: Sensory processes and perception* (pp. 10–102). New York: Wiley.
Hanna, J. E., Tanenhaus, M. K., & Trueswell, J. C. (2003). The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language, 49*(1), 43–61.
Harris, J. A., & Potts, C. (2009). Perspective-shifting with appositives and expressives. *Linguistics and Philosophy, 32*, 523–552.
Heller, D., & Brown-Schmidt, S. (2023). The multiple perspectives theory of mental states in communication. *Cognitive Science, 47*(7), Article e13322.
Heller, D., Parisien, C., & Stevenson, S. (2016). Perspective-taking behavior as the probabilistic weighing of multiple domains. *Cognition, 149*, 104–120.
Horton, W. S. (2005). Conversational common ground and memory processes in language production. *Discourse Processes, 40*(1), 1–35.
Horton, W. S., & Gerrig, R. J. (2002). Speakers' experiences and audience design: Knowing when and knowing how to adjust utterances to addressees. *Journal of Memory and Language, 47*(4), 589–606.
Horton, W. S., & Gerrig, R. J. (2016). Revisiting the memory-based processing approach to common ground. *Topics in Cognitive Science, 8*(4), 780–795.
Jara-Ettinger, J., & Rubio-Fernandez, P. (2022). The social basis of referential communication: Speakers construct physical reference based on listeners' expected visual search. *Psychological Review, 129*(6), 1394.
Kangasharju, H. (1996). Aligning as a team in multiparty conversation. *Journal of Pragmatics, 26*(3), 291–319.
Kärkkäinen, E. (2006). Stance taking in conversation: From subjectivity to intersubjectivity. *Text & Talk (Print), 26*(6), 699–731.
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.

Keysar, B., Barr, D. J., Balin, J. A., & Paek, T. S. (1998). Definite reference and mutual knowledge: Process models of common ground in comprehension. *Journal of Memory and Language, 39*(1), 1–20.
Krauss, R. M., & Fussell, S. R. (1991). Perspective-taking in communication: Representations of others' knowledge in reference. *Social Cognition, 9*(1), 2–24.
Li, C. N., & Thompson, S. (1975). The semantic function of word order: A case study in Mandarin. In *Word order and word order change* (pp. 163–186). University of Texas Press.
Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. University of California Press.
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.
Mozuraitis, M., Chambers, C. G., & Daneman, M. (2015). Privileged versus shared knowledge about object identity in real-time referential processing. *Cognition, 142*, 148–165.
Mozuraitis, M., Stevenson, S., & Heller, D. (2018). Modeling reference production as the probabilistic combination of multiple perspectives. *Cognitive Science, 42*, 974–1008.
Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(2), 169–190.
Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–347.
Pickering, M. J., & Garrod, S. (2021). *Understanding dialogue: Language use and social interaction*. Cambridge University Press.
Ritchie, D. (2004). Common ground in metaphor theory: Continuing the conversation. *Metaphor and Symbol, 19*(3), 233–244.
Robbins, J. M., & Krueger, J. I. (2005). Social projection to ingroups and outgroups: A review and meta-analysis. *Personality and Social Psychology Review, 9*(1), 32–47.
Savitsky, K., Keysar, B., Epley, N., Carter, T., & Swanson, A. (2011). The closeness-communication bias: Increased egocentrism among friends versus strangers. *Journal of Experimental Social Psychology, 47*(1), 269–273.
Sperber, D., & Wilson, D. (1986). *Relevance: Communication and cognition*. Harvard University Press.
Stalnaker, R. (2002). Common ground. *Linguistics and Philosophy, 25*(5/6), 701–721.
Wu, S., & Keysar, B. (2007). The effect of information overlap on communication effectiveness. *Cognitive Science, 31*(1), 169–181.
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.
Yoon, S. O., & Brown-Schmidt, S. (2018). Aim low: Mechanisms of audience design in multiparty conversation. *Discourse Processes, 55*(7), 566–592.
Yoon, S. O., & Brown-Schmidt, S. (2019). Audience design in multiparty conversation. *Cognitive Science, 43*(8), Article e12774.
Zheng, X., & Breheny, R. (2021). Investigating the basis of memory-based effects on common ground. *Language, Cognition and Neuroscience, 36*(8), 941–958.
Zheng, X., & Sun, C. (2025). Understanding perspective-taking in multiparty conversations: Insights from Mandarin nouns. *Frontiers in Psychology, 16*, Article 1499538.