Discourse cues for the acquisition of the Mandarin contrafactive verb *yǐwéi*

Dingyi Pan (dipan@ucsd.edu)
Department of Linguistics, UC San Diego
La Jolla, CA, 92093

Rachel Dudley (radudley@ucsd.edu)
Department of Linguistics, UC San Diego
La Jolla, CA, 92093

Abstract

In contrast to nonfactive belief verbs, such as English "think" and Mandarin *juéde* (to think/feel), the Mandarin contrafactive belief verb *yǐwéi* is negatively biased and commonly used to report false beliefs (Glass, 2023, 2025). Previous studies show that children can distinguish between nonfactives and contrafactives by 4 years of age, and the use of *yǐwéi* facilitates understanding of false-belief sentences and false-belief reasoning more generally (Lee, Olson, & Torrance, 1999; Zhang & Zhou, 2022). However, it remains unclear how children learn that *yǐwéi* is contrafactive. One possible factor is the discourse context in which *yǐwéi* occurs: compared to *juéde*, *yǐwéi* is more likely to occur in contexts that highlight that the content of the attributed belief is false. In this study, we investigate whether this kind of discourse information can help Mandarin-speaking children acquire the contrafactive verb *yǐwéi* using a modified version of the Human Simulation Paradigm (Gillette, Gleitman, Gleitman, & Lederer, 1999). The results suggest that discourse might be one of the cues facilitating the acquisition of *yǐwéi*, while other linguistic cues remain to be explored.

Keywords: contrafactives; language acquisition; Human Simulation Paradigm

Introduction

Different mental state verbs express different relationships between the attitude holder, which is the subject in the matrix clause, and the embedded proposition, which is expressed in the embedded clause. Factives, such as "know p", both entail and presuppose the truth of their embedded content *p*. Thus, "know" can only be used to report true beliefs, and upon hearing utterances like (1-a), the listener can infer that the speaker takes it for granted that *it is raining outside*. In contrast, nonfactives, like "think p" in (1-b), simply assert that the agent holds the belief that *it is raining*, but it can be the case that it is not raining outside and John holds a false belief. Thus, "think" can be used to report either true beliefs or false beliefs, and the listener will not be able to infer whether *it is raining outside*.

(1) a. John knows that it is raining outside.
b. John thinks that it is raining outside.

In English, factive verbs like "know," which are restricted to attributions of true belief (or knowledge), have no counterparts that are restricted to attributions of false beliefs. Such a contrafactive verb, such as "CONTRA," would be useful to explicitly signal that the attitude holder is mistaken and their belief is false. Although nonfactives like "think" can also be used to report false beliefs, they are not restricted to such contexts: nonfactives may be used when *p* is either true or false. As a result, while "think p" leaves the truth value of *p* undetermined, "CONTRA p" would be more informative and allow the listener to infer straightforwardly that *p* is false.

Mandarin presents a case that may fill this lexical gap for contrafactives: *yǐwéi*. For instance, in (2), we not only take the subject, Xiaoming, to believe in the embedded proposition *p*, but we also infer that *p* is false, i.e., *it was not raining outside*.[1]

(2) xiǎomíng yǐwéi wàimian xiàyǔ le.
xiaoming yǐwéi outside rain ASP.
'Xiaoming falsely believed that it was raining outside.'

This contrafactive is highly useful, especially in facilitating children's understanding of false beliefs. As compared to test questions involving the nonfactive verb *xiǎng* (to think), Mandarin-speaking children were more likely to correctly report false beliefs when *yǐwéi* was used in the test question (Lee et al., 1999). In the same study by Lee et al. (1999), Mandarin-speaking adults also preferred *yǐwéi* over the nonfactive *xiǎng* when describing false-belief scenarios.

Although contrafactives are more informative and helpful for highlighting false beliefs, Holton (2017) argues that no Indo-European language has an attested contrafactive that is both a non-decomposable lexical unit and presupposes the falsity of the embedded content. Various hypotheses have been proposed to explain the apparent lexical gap for contrafactives. One of them is the learnability account, which holds that contrafactives are more difficult to learn than factives. Results from a study using an artificial learning paradigm with English-speaking adults (Maldonado, Culbertson, & Uegaki, 2022) and from computational simulations (Strohmaier & Wimmer, 2023, 2024, 2025) support this learning constraint. Yet, in contrast to learnability studies in Indo-European languages, native Mandarin speakers have no difficulty using and understanding *yǐwéi*, and Mandarin-speaking children can interpret the negative bias of *yǐwéi* at

[1] We consider *yǐwéi* to be a contrafactive by adopting the logically weaker definition of contrafactives proposed in Glass (2025). Specifically, it only requires the compatibility of the Common Ground with *not p*, and only after the assertion containing the contrafactive is accepted into the Common Ground.

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

the age of 4 (Zhang & Zhou, 2022). Given the mismatch between the high utility of contrafactives and their crosslinguistic rarity, how do Mandarin-speaking children learn this contrafactive verb? In this study, we show that discourse contexts provide useful learning cues that constrain the meaning space, while sentential cues might further distinguish *yǐwéi* from nonfactive verbs when the discourse is uninformative.

Developmental studies with Mandarin *yǐwéi*

To date, most studies involving *yǐwéi* have focused on its effect on children's false-belief reasoning. For instance, Lee et al. (1999) showed that children aged 3-5 performed better on false-belief tasks when *yǐwéi* was used as part of the dependent measure. In addition, children's comprehension of *yǐwéi* correlates with their performance in false-belief tasks between the ages of 4 and 5 (Brandt, Li, & Chan, 2023). The existence of *yǐwéi* is commonly believed to help Mandarin-speaking children succeed in false-belief tasks, as this verb contrasts with nonfactive verbs and thus highlights that the attributed belief may be false. In a recent study investigating children's understanding of belief-reporting sentences, Zhang and Zhou (2022) found that both Mandarin-speaking adults and 4-year-olds are more likely to interpret *p* as false when it is embedded under *yǐwéi* than under *juéde* (to think/feel). Importantly, children showed better understanding of false-belief attributions with *yǐwéi* as the matrix verb than with *juéde*. Nonetheless, it remains unclear how *yǐwéi* is learned.

Acquisition of mental verbs

Mental state verbs are challenging for children to learn. While some argue that this difficulty stems from children's conceptual constraints, others suggest that mental state verbs do not provide physically observable referents and, thus, that the mapping problem is more difficult. According to the latter view, there may be multiple sources of information needed to learn words, and children could rely on linguistic sources of information, in addition to extralinguistic ones. One common strategy that attempts to isolate the conceptual challenge and focus on the information sources during word learning is the Human Simulation Paradigm (Gillette et al., 1999). In the Human Simulation Paradigm, adults, who have mature language and conceptual systems and thus should not have conceptual constraints, are presented with a new word to simulate the learning problem faced by children. Various linguistic and extralinguistic cues can be ablated or provided to the participant to support inferences about which sources of information children might rely on during word learning.

In the case of mental verbs, even adults struggle to correctly guess mental state verbs when only extralinguistic information is present (Gillette et al., 1999; Snedeker & Gleitman, 2004; Gleitman, Cassidy, Nappa, Papafragou, & Trueswell, 2005, a.o.). Because mental state verbs are more abstract than concrete action verbs, and because verbs work as linguistic scaffolding, they benefit from their syntactic distribution during word learning, a mechanism commonly referred to as syntactic bootstrapping (Gleitman, 1990, a.o.). Moreover, while syntactic structure helps constrain the plausible interpretation of a novel word to mentalistic meanings, adding extralinguistic cues that highlight the salience of mental states, such as the presentation of a false-belief scenario, can further prompt reference to mental states (Papafragou, Cassidy, & Gleitman, 2007).

Contextual cues and discourse bootstrapping

Beyond the domain of mental state verb acquisition, studies have shown that children are sensitive to contextual information, and contextually salient events can facilitate children's learning of abstract terms. A recent study by Gomes, Huh, Yun, and Trueswell (2026) suggests that failures and violations of expectations can be "gems" to the acquisition of terms for truth-functional negation, such as "not" in English, since these cues may be rare and infrequent but transparent to the learner and extremely effective in highlighting negative meaning (Gleitman & Trueswell, 2020). This is because failures may be more valuable to mention than successful completions, and adults are also more likely to use negators to comment on them. Thus, the salience of failures may be particularly informative for learning the meaning of negators, in addition to linguistic cues.

Moreover, children are also sensitive to linguistically encoded contextual information and use discourse information during word learning. Sullivan and Barner (2016) provides support for the hypothesis that children use discourse information to "bootstrap" themselves into the meaning of novel words. In a similar study, children around the age of 4 can rely on discourse relations, as indicated by discourse connectives, to infer the meanings of new words, further supporting the effect of discourse on verb learning (Sullivan, Boucher, Kiefer, Williams, & Barner, 2019). Work on syntactic bootstrapping also suggests that pragmatic information is leveraged in addition to syntactic information to help the learner further constrain the hypothesis space (Hacquard & Lidz, 2019; Fisher, Jin, & Scott, 2020).

With this study, we pursue the idea that, similar to how failures are "gems" for the learning of negators, discourse contexts that clearly provide information which contradicts the embedded content *p* might be one of the "gems" for the acquisition of *yǐwéi*. When false beliefs are salient in the context, parents may be more likely to comment on them, and *yǐwéi* is especially informative for such comments because, unlike nonfactive verbs, it unambiguously conveys that *p* cannot be true.

Current Study

Motivated by findings on the use of contextual information in word learning and the discourse bootstrapping hypothesis, we hypothesize that discourse information which contrasts with the embedded content *p* can facilitate the learning of the Mandarin contrafactive *yǐwéi p*. Although both nonfactives and contrafactives are compatible in situations where *p* is false in the discourse context, *yǐwéi* is usually preferred over the neutral alternative *juéde* because it is more informative, as

suggested by the findings in Lee et al. (1999).[2] To test this hypothesis, we adopt a modified version of the Human Simulation Paradigm. In Experiment 1, we present participants with dialogues between adults and children in which the verbs *yǐwéi* and *juéde* are removed from the target sentences. Participants are asked to recover the missing verb based on the discourse context and the rest of the target sentence. In Experiment 2, we present only the target sentence without its discourse context, in order to isolate the effect of discourse information from other linguistic cues.[3]

Experiment 1: In the presence of contexts

Methods

Materials As part of an ongoing corpus study, we first extracted child-directed speech that contained the two target verbs and their surrounding context from the Mandarin CHILDES corpora (Macwhinney, 2000), and *juéde* is 24 times more frequent than *yǐwéi*. We limited the context presented to the participants to the five lines preceding the target sentence and the two lines following it, since we assume the local context is more salient to interlocutors and our coding suggests that discourse cues usually occur in the immediate context of the target sentence. Hence, each dialogue snippet contained eight lines.

Two independent coders, who were linguistically trained and native speakers of Mandarin, coded whether the discourse context provides explicit information about *p*. When the discourse provides explicit information about *p*, we further coded whether it supports *p* or *not p*. For a given embedded content *p* (e.g., "It is a cat."), we considered the context to be supporting *not p* in two cases: First, when *not p* is explicitly mentioned in the context (e.g., "It is not a cat."), and second, when the context contains some alternative proposition *q* that cannot be true at the same time as *p* (e.g., "It is a dog."). Whether a discourse provides information about the truth of *p*, we used it in the "*constrained*" condition. If the discourse provided no restricting information about *p*, that is, supporting neither *p* nor *not p*, we used it in the "*unconstrained*" condition. Both types of contexts were considered for each verb. In the *constrained* condition, we only included dialogues where *p* is supported in the discourse context for *juéde* and *not p* is supported for *yǐwéi*. Note that semantically speaking, *juéde* does not require the context to provide any information about *p* and is compatible in both types of *constrained* contexts, where it can also be used in the *not p* contexts selected for *yǐwéi*. In contrast, *yǐwéi* cannot be used in the contexts selected for *juéde* that support *p*, given the incompatibility of *yǐwéi p* with *p*.

Thus, the combination of these two verbs and two types of discourse cues yields four critical conditions. The English translations of the example dialogues are provided in Table 1. Forty test items were drawn from the larger dataset in the corpus study, with 10 dialogues in each condition.[4] Two coders agreed on 39 out of 40 items about the type of discourse cues, either *constrained* (supporting *p* or *not p*) or *unconstrained*. An additional 10 dialogues containing either Chinese four-character idioms or adverbial temporal phrases, *yǐqián* (before/previously) and *yǐhòu* (after/later), were also included as controls to assess participants' language proficiency.

For critical items, we removed the target verb from the target sentence and asked participants to select between the two options, *yǐwéi* or *juéde*, in a two-alternative forced-choice task. For the control items, we similarly removed the adverbial temporal phrase or two characters from the four-character idioms, and participants were instructed to select between two candidates.

Procedure Participants were instructed that a word was left blank in a dialogue between parents (or other adults) and a child, and their task was to choose between two options corresponding to what they thought the adult might have used. The experiment started with four practice trials using the same task format to familiarize the participant with the task. Explicit feedback was given at the end of each practice trial. Then each participant saw all 50 critical dialogues, with order randomized across participants.

Participants We recruited 50 Mandarin-speaking participants on Prolific. Results from 8 participants were excluded based on the preregistered exclusion criteria: 7 participants did not self-report as native Mandarin speakers, and 1 had an accuracy of 80% or less on the control items. Results from the remaining 42 participants were included in the analysis.

Predictions

We considered an answer to be correct if it matched the verb use in the original corpus. If discourse cues facilitate the acquisition of *yǐwéi*, Mandarin-speaking adults should be more likely to accurately recover *yǐwéi* than to incorrectly guess *juéde* when the discourse clearly supports *not p*. In contrast, although both verbs can be used when the discourse provides no information about *p*, participants may be more likely to choose *juéde* given its substantially higher frequency compared to *yǐwéi*, leading to lower accuracy in the *yǐwéi* + *unconstrained* condition and higher accuracy in the *juéde* + *constrained* condition. Lastly, the accuracy of *juéde* may be high in the *constrained* condition, since only *juéde* is compatible with the discourse that supports *p*.

Results

Figure 1(a) shows the mean accuracy in participants' guesses of the target verb. The mean accuracy across the four conditions was above chance, with *juéde* + *unconstrained* and

[2] *juéde* is used as the nonfactive belief verb instead of the *rènwéi* because *rènwéi* is extremely infrequent in the CHILDES datasets with only 9 occurrences in child-directed speech. In addition, *juéde* can also mean "to feel" in Mandarin, but we focus on cases where *juéde* is used as a belief-reporting verb that embeds a syntactic object which expresses a proposition.

[3] The pre-registration is available at `https://osf.io/hn7vx`. All data, materials, and analysis scripts can be accessed at `https://github.com/pennydy/yiwei`.

[4] Data for one dialogue in the *yǐwéi* + *constrained* condition were excluded in Experiment 1 due to experimenter error.

Table 1: English translation for example dialogues in each condition. The target verbs, *yǐwéi* and *juéde*, are underlined.

	Constrained discourse with *p/not p*	**Unconstrained discourse**
yǐwéi	ADULT: This was really well done. ADULT: Who made this? CHILD: (This was) made by Uncle. ADULT: Oh! ADULT: It was Uncle who made it? ADULT: I *yǐwéi* it was made by Grandma. ADULT: Such a beautiful car. ADULT: Oh no!	CHILD: The kitten also wanted to play outside. MOTHER: Look, this owner put up a missing cat poster. MOTHER: Let me read it to you. Lost cat notice, okay? CHILD: Okay. MOTHER: You disappeared. MOTHER: I *yǐwéi* it was just our usual hide-and-seek game. MOTHER: You must be hiding in some corner, secretly laughing at me. MOTHER: This is the fifth day, you've been hiding for so long.
juéde	MOTHER: Is it a mosquito? CHILD: No. MOTHER: Then what is it? CHILD: A ladybug. MOTHER: Oh, it is a ladybug. MOTHER: You *juéde* it is a ladybug. MOTHER: The spider didn't respond. MOTHER: He's busy spinning the web.	MOTHER: Look, doesn't it look like an egg? MOTHER: How to put it together? CHILD: Put it together like this! MOTHER: Put it together like this. MOTHER: Right. MOTHER: I also *juéde* this puzzle is a little hard. MOTHER: Okay. MOTHER: Ah, I know.

yǐwéi + *constrained* near-ceiling, and *yǐwéi* + *unconstrained* lower than in the other three conditions. In addition, there was more variability in participants' performance in the *unsupported* discourse condition than in the *constrained* discourse condition, especially when the verb was *yǐwéi*. A logistic mixed-effect regression was conducted to predict the accuracy of participants' guesses from sum-coded fixed effects of verb type, discourse type, and their interaction. Also included in the model are the random by-item intercept and the random by-participant intercept and slopes for the fixed effects and their interaction.[5] The main effect of verb type was marginally significant, with the accuracy higher for *juéde* than for *yǐwéi* ($\beta = 0.53$, $SE = 0.28$, $z = 1.10$, $p = 0.0587$). However, there was no significant main effect of discourse type ($\beta = 0.11$, $SE = 0.26$, $z = 0.42$, $p = 0.6764$), and its interaction with verb type also did not reach significance ($\beta = -0.36$, $SE = 0.26$, $z = -1.40$, $p = 0.1627$). As part of the preregistered analysis, planned pairwise comparisons were conducted on the fitted logistic mixed-effect model using estimated marginal means, although the interaction was not significant.[6] In the *constrained* conditions, the accuracy did not differ significantly between the two verbs ($\beta = 0.34$, $SE = 0.75$, $z = 0.44$, $p = 0.6558$). Yet, in the *unconstrained* conditions, *juéde* had higher accuracy than *yǐwéi* ($\beta = 1.79$, $SE = 0.78$, $z = 2.30$, $p = 0.0213$).

Discussion

The relatively low accuracy of *yǐwéi* in the *unconstrained* discourse condition trends in the predicted direction, suggesting that *yǐwéi* might rely more on the discourse information than *juéde*, despite limited statistical significance. The ability to detect differences between conditions might be due to a ceiling effect, as indicated by the overall high accuracy. One possible reason for this ceiling effect is that the two-alternative forced-choice experimental setup raised the likelihood of participants using the relatively infrequent verb *yǐwéi*.

In addition, 10 items per condition may be insufficient, and the analysis may have limited power at the item level. The random-effects estimates revealed substantial item-level variability ($SD = 1.37$ on the log-odds scale), substantially larger than the participant-level variability ($SD = 0.76$), suggesting considerable variability across items that might reduce sensitivity to main fixed effects. Thus, as an exploratory analysis, we fit an additional model identical to the one reported in the Results section but without the random by-item intercept. There were significant effects of verb type ($\beta = 0.30$, $SE = 0.12$, $z = 2.48$, $p = 0.0131$) and discourse type ($\beta = 0.18$, $SE = 0.09$, $z = 2.01$, $p = 0.0449$) as well as their interaction ($\beta = -0.33$, $SE = 0.08$, $z = -3.90$, $p < 0.001$). We return to this issue in General Discussion.

Setting aside the suggestive pattern, it is also possible that the discourse context is not the only information that participants used to recover the verbs. Syntactic or lexical information in the target sentence itself, such as the presence of certain adverbs or the co-occurrence of the matrix and embedded subjects, might have provided cues to the correct verb choice. Although Mandarin is morphosyntactically impoverished, previous work has found that syntactic signals are available to distinguish between different lexical classes, such as belief and desire verbs (Huang, White, Liao, Hacquard, & Lidz, 2022). Therefore, similar syntactic cues may be available to separate *yǐwéi* from *juéde*. In Experiment 2, we removed context sentences and presented participants only with the target sentence to isolate the effect of discourse.

Experiment 2: In the absence of contexts

Methods

Materials and procedure In this experiment, we ablated all discourse information in order to determine how much signal remains when presented with the syntactic and lexical information in the target sentence. Only this target sentence,

[5] All regression models reported in this paper were run using the `glmer` package (Bates, Mächler, Bolker, & Walker, 2015).

[6] All pairwise comparisons of marginal means reported in this paper were estimated using `emmeans` package (Lenth, 2025).

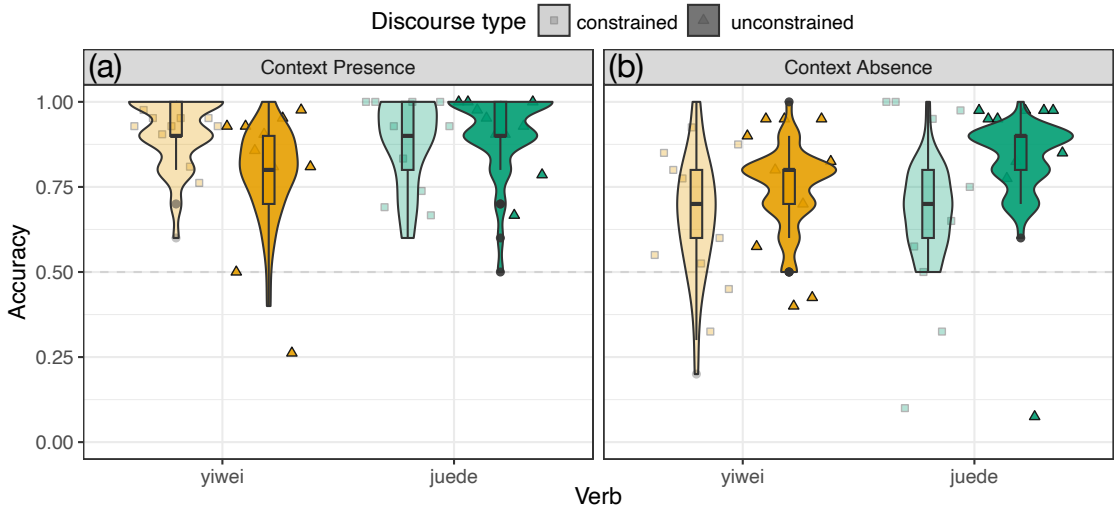

Figure 1: Mean accuracy of participants' guesses by verb and discourse types with discourse contexts (a) and without discourse contexts (b). The colored triangles and squares represent the mean accuracy for each item.

extracted from the original dialogue, was presented to a separate group of participants, allowing us to further disentangle the role of discourse from sentence-internal cues. Hence, the same set of 50 items as Experiment 1 was used, and we followed the same procedure as Experiment 1.

Participants Fifty participants who did not participate in Experiment 1 were recruited on Prolific. Results from 10 participants were excluded using the same preregistered exclusion criteria as in Experiment 1: 7 participants did not self-report as native Mandarin speakers, and 3 had less than or equal to 80% accuracy on the control items. Results from the remaining 40 participants were included in the final analysis.

Results

The mean accuracy of the participants' guesses of the target verb is shown in Figure 1(b). While the accuracy was above chance across all four conditions, the accuracy was higher in *juéde* + *unconstrained* than that in the other three conditions. In addition, substantial variability between items was observed in all four conditions. Using the same analysis as in Experiment 1, we fit a logistic mixed-effects regression to predict participants' guess accuracy from the sum-coded main effects of verb type, discourse type, and their interaction, as well as the random by-item intercept and random by-participant intercept and slopes for verb type, discourse type, and their interaction. Neither of the two main effects (Verb type: $\beta = 0.32$, $SE = 0.27$, $z = 1.19$, $p = 0.235$; Discourse type: $\beta = -0.36$, $SE = 0.27$, $z = -1.33$, $p = 0.184$) nor their interaction was significant ($\beta = -0.09$, $SE = 0.27$, $z = -0.33$, $p = 0.742$).

As in Experiment 1, this analysis may be limited by low power at the item level, given the small number of items per condition. Therefore, as an exploratory analysis comparable to the one conducted in Experiment 1, we fit a model without the by-item random intercept. There was a main effect of verb, where *juéde* had higher accuracy than *yǐwéi* ($\beta = 0.14$, $SE = 0.06$, $z = 2.24$, $p = 0.0252$). The effect of discourse type was also significant ($\beta = -0.30$, $SE = 0.06$, $z = -4.84$, $p < 0.001$), and its interaction with verb was marginally significant ($\beta = 0.12$, $SE = 0.06$, $z = -1.91$, $p = 0.0556$).

Comparison with Exp.1 As part of the preregistered analysis, we compared results between the two Experiments, with and without discourse context. In general, participants were less likely to guess the correct verb in the absence of the discourse context than in the presence of the discourse context. This decrease in accuracy was more pronounced in the *constrained* than in the *unconstrained* discourse conditions, especially for *yǐwéi*: The absence of constrained discourse context that supports *not p* led to lower accuracy of correctly guessing *yǐwéi*, whereas the effect is less so when the discourse context is not constrained.

These observations were borne out statistically. We fit a logistic mixed-effect model, predicting participants' accuracy from sum-coded main effects of verb type, discourse type, context presence, and their interaction, as well as the maximal random effects structure that allowed the model to converge (random by-item intercept and slope for context presence and random by-participant intercept and slopes for verb type, discourse type, and context presence). There was a marginally significant effect of verb type ($\beta = 0.46$, $SE = 0.25$, $z = 1.84$, $p = 0.0656$), suggesting that *juéde* had higher accuracy than *yǐwéi*, which might be due to its higher overall frequency. In addition, the effect of discourse context presence was significant ($\beta = -0.61$, $SE = 0.12$, $z = -5.27$, $p < 0.001$), suggesting that participants' accuracy was higher in the presence of the context than in its absence. In addition, the interaction between discourse type and context presence was also significant ($\beta = -0.28$, $SE = 0.10$, $z = -2.81$, $p = 0.00502$), where the effect of discourse type was smaller in the absence of context than in its presence. Pairwise comparisons showed that,

when the discourse supported *p* or *not p*, accuracy was much higher when participants saw the discourse context than when they did not ($\beta = -1.79$, $SE = 0.31$, $z = -5.75$, $p < 0.001$). The same relationship held in the discourse *unconstrained* conditions ($\beta = -0.68$, $SE = 0.30$, $z = -2.25$, $p = 0.0243$). Other main effects and interaction terms were not significant.

Discussion

Although the accuracy across all four conditions in Experiment 2 remained above chance, it was lower than in Experiment 1 with the discourse context present. In addition, the effect of the absence of the discourse context on accuracy was larger in the *constrained* discourse condition than in the *unconstrained* discourse condition, suggesting that discourse context is most useful in recovering the verb when it provides information about *p*.

We might have expected accuracy to drop below chance in Experiment 2 if only discourse cues mattered in distinguishing *juéde* from *yǐwéi*. However, given that adults still achieved high accuracy in recovering the correct verb in the absence of discourse contexts, this raises the possibility that both discourse cues and other sentence-level cues contribute to distinguishing between the verbs. We even see the suggestion of a trade-off between the two: While discourse cues are the strongest indicators of the verb choice, people might rely on the syntactic structure or lexical co-occurrence when such cues are not available, resulting in relatively high accuracy in the *unconstrained* and context-absent conditions. Therefore, it is possible that discourse is important for identifying *yǐwéi* when it supports the constraint. Yet, when the discourse context is not compatible with the constraint or is absent, the sentential cues are enough to distinguish between the two verbs.

General discussion

In this study, we investigated how adults recover a missing belief verb in Mandarin in order to provide insight into the factors that facilitate children's acquisition of the contrafactive *yǐwéi*. Inspired by discourse bootstrapping in reference identification and the use of contextually salient failures in the acquisition of negators, we hypothesized that discourse contexts that provide information about the discourse status of the attributed belief can serve as effective cues, since false beliefs may be salient and therefore more likely to elicit comments from parents. The results show that, compared to when the discourse cue is absent or does not provide information about *p*, having discourse support for *not p* improved the accuracy of participants' guesses for *yǐwéi*, above and beyond any improvement in accuracy for *juéde*.

In addition, the results of the current study can potentially inform the learnability account of contrafactives. Specifically, if Mandarin-speaking children rely on discourse cues to constrain the meaning space of *yǐwéi*, will computational models similar to those trained in Strohmaier and Wimmer (2025) or adult speakers of a language without *yǐwéi*-like contrafactives, as in Maldonado et al. (2022), learn the verb when provided the same types of cues?

Despite these suggestive results and implications, there are a few limitations of the current study. As evident in both experiments, accuracy was high across the board. One possible reason for this ceiling effect is the choice of the dependent measure. Since *yǐwéi* is far less frequent than *juéde*, having *yǐwéi* as an option in a two-alternative forced-choice experimental setup might increase its usage by participants. Future studies can use an open-ended dependent measure, requiring participants to fill out the missing verb, similar to the original Human Simulation Paradigm design. Participants might then generate a wider range of possible responses, including other nonfactive verbs like *rènwéi* (to think) and factive verbs like *zhīdào* (to know). It would be informative to examine whether *yǐwéi* remains relatively preferred when discourse supports *not p* once the potential advantage from the forced-choice design is removed.

Moreover, as argued in prior literature, multiple sources of information can serve as cues for hypothesizing the meanings of new attitude verbs (Gleitman et al., 2005; Papafragou et al., 2007), and discourse information constitutes only one of the many sources available to children. In the case of *yǐwéi*, these results suggest that the discourse context is useful to the learner, but that other cues may also facilitate learning. As shown by the comparison between Experiments 1 and 2, the discourse context might help constrain the possible meaning space for *yǐwéi*, and when it is not available, as in the *unconstrained* and context absence conditions, people might rely on sentential cues. As a preliminary step towards identifying potential sentence-level cues, we analyzed the 40 target sentences used in this study. 12 of the 20 sentences with *yǐwéi* had a first-person pronoun as the matrix subject, and all reported their own false beliefs about others. In contrast, *juéde* had 8 of 20 sentences with a first-person pronoun as the matrix subject, and two were cases where the speaker reported a belief about themself. In addition, *yǐwéi* occurred only once with a second-person subject in the discourse *unconstrained* context, much less frequently than *juéde*. While this is limited to the 40 sentences used in this study, an ongoing corpus study aims to analyze a wider range of information, including the pairing between certain modifiers (e.g., *běn/yuán (lái)*, originally) and *yǐwéi* and the use of different matrix and embedded subjects with *yǐwéi* for different discourse functions. Armed with more information about the distribution of *yǐwéi* vs. *juéde* in naturalistic speech, we can use the same Human Simulation Paradigm to individually control and vary these variables to infer their impact on the acquisition of *yǐwéi*.

In addition, because our goal was to identify cues in children's learning inputs that could serve as "gems" for the acquisition of *yǐwéi*, we considered only child-directed speech sampled from the CHILDES datasets. Thus, we were limited to 10 items per condition, resulting in limited statistical power at the item level. One possible way to alleviate this issue is to artificially generate more stimuli. Future studies can explore this strategy or use additional adult-directed speech corpora to scale up the stimulus set.

Acknowledgments

We would like to thank Yage Grace Xin and Hongao Zhu for their help with the discourse coding. We are also grateful to members of the Meaning & Language Development Lab at UCSD and participants at the second Southern California Meeting for Investigations in Developmental Science for helpful comments and discussion.

References

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

Brandt, S., Li, H., & Chan, A. (2023). What makes a complement false? looking at the effects of verbal semantics and perspective in mandarin children's interpretation of complement-clause constructions and their false-belief understanding. *Cognitive Linguistics*, *34*(1), 99–132.

Fisher, C., Jin, K.-s., & Scott, R. M. (2020). The developmental origins of syntactic bootstrapping. *Topics in Cognitive Science*, *12*(1), 48–77.

Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, *73*(2), 135-176. doi: https://doi.org/10.1016/S0010-0277(99)00036-0

Glass, L. (2023). The negatively biased mandarin belief verb yǐwéi. *Studia Linguistica*, *77*(1), 1-46.

Glass, L. (2025, July). Attested versus unattested contrafactive belief verbs. *Semantics and Pragmatics*, *18*(8), 1–17. doi: 10.3765/sp.18.8

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55. doi: 10.1207/s15327817la0101_2

Gleitman, L. R., Cassidy, K., Nappa, R., Papafragou, A., & Trueswell, J. C. (2005). Hard words. *Language learning and development*, *1*(1), 23–64.

Gleitman, L. R., & Trueswell, J. C. (2020). Easy words: Reference resolution in a malevolent referent world. *Topics in Cognitive Science*, *12*(1), 22-47. doi: https://doi.org/10.1111/tops.12352

Gomes, V., Huh, Y., Yun, H., & Trueswell, J. C. (2026). Failure is always an option: Children's production of truth-functional negation and its implications for learning. *Cognition*, *267*, 106342. doi: https://doi.org/10.1016/j.cognition.2025.106342

Hacquard, V., & Lidz, J. (2019). Children's attitude problems: Bootstrapping verb meaning from syntax and pragmatics. *Mind & Language*, *34*(1), 73–96.

Holton, R. (2017). Facts, factives, and contrafactives. *Aristotelian Society Supplementary Volume*, *91*(1), 245–266. doi: 10.1093/arisup/akx003

Huang, N., White, A. S., Liao, C.-H., Hacquard, V., & Lidz, J. (2022). Syntactic bootstrapping attitude verbs despite impoverished morphosyntax. *Language Acquisition*, *29*(1), 27–53.

Lee, K., Olson, D. R., & Torrance, N. (1999). Chinese children's understanding of false beliefs: The role of language. *Journal of Child Language*, *26*(1), 1–21.

Lenth, R. V. (2025). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. (R package version 1.11.1) doi: 10.32614/CRAN.package.emmeans

Macwhinney, B. (2000, 01). The CHILDES project: tools for analyzing talk. *Child Language Teaching and Therapy*, *8*. doi: 10.1177/026565909200800211

Maldonado, M., Culbertson, J., & Uegaki, W. (2022). Learnability and constraints on the semantics of clause-embedding predicates. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *44*(44).

Papafragou, A., Cassidy, K., & Gleitman, L. R. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition*, *105*(1), 125-165. doi: https://doi.org/10.1016/j.cognition.2006.09.008

Snedeker, J., & Gleitman, L. (2004). Why it is hard to label our concepts. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon* (pp. 257–294). Cambridge, MA: MIT Press. doi: 10.7551/mitpress/7185.001.0001

Strohmaier, D., & Wimmer, S. (2023). Contrafactives and Learnability: An Experiment with Propositional Constants. In D. Bekki, K. Mineshima, & E. McCready (Eds.), *Logic and Engineering of Natural Language Semantics* (pp. 67–82). Cham: Springer Nature Switzerland. doi: 10.1007/978-3-031-43977-3_5

Strohmaier, D., & Wimmer, S. (2024). Contrafactives and learnability. In F. Carcassi, T. Johnson, S. B. Knudstorp, S. D. Parrado, P. R. Robledo, & G. Sbardolini (Eds.), *Proceedings of the 24th amsterdam colloquium* (pp. 298–305).

Strohmaier, D., & Wimmer, S. (2025). Contrafactives, learnability, and production. *Experiments in Linguistic Meaning*, *3*, 395–410. doi: 10.3765/elm.3.5810

Sullivan, J., & Barner, D. (2016). Discourse bootstrapping: preschoolers use linguistic discourse to learn new words. *Developmental Science*, *19*(1), 63-75. doi: https://doi.org/10.1111/desc.12289

Sullivan, J., Boucher, J., Kiefer, R. J., Williams, K., & Barner, D. (2019). Discourse coherence as a cue to reference in word learning: Evidence for discourse bootstrapping. *Cognitive Science*, *43*(1), e12702. doi: https://doi.org/10.1111/cogs.12702

Zhang, X., & Zhou, P. (2022). Linguistic cues facilitate children's understanding of belief-reporting sentences. *First Language*, *42*(1), 51-80. doi: 10.1177/01427237211048669