When Parsing and interpretation misalign: a case of wh-scope ambiguity resolution in Mandarin

Abstract

A great amount of sentence processing work has focused on revealing how the parser incrementally integrates each incoming word into the current linguistic representation. It is often explicitly or implicitly assumed that the representation preferred by the parser would determine the ultimate interpretation of the sentence. The current study investigates whether the interpretive bias in sentence comprehension necessarily tracks the parsing bias. Our case study is concerned with the locality bias in non-local dependencies, specifically, the Mandarin wh-in-situ scope dependencies. Our findings suggest a misalignment between the local parsing decisions and the global interpretative decisions. In particular, for Mandarin wh-in-situ constructions that involve scope ambiguity, there is a locality bias in parsing, but there is an anti-locality bias in interpretation. We propose a bayesian pragmatic inference analysis to account for these findings. Under this proposal, the seeming conflict between parsing and interpretation will ultimately disappear because parsing preferences will be naturally embedded under the pragmatic reasoning process to derive the ultimate interpretation. The current study therefore makes novel contributions, both empirically and theoretically, to address questions about the relationship between parsing and interpretation.

Keywords:

1. Introduction

Sentence comprehension requires a parser that establishes the structural representation of the to-be-interpreted sentence. A great amount of sentence processing work has focused on revealing how the parser incrementally integrates each incoming word into the current linguistic representation. It is often explicitly or implicitly assumed that the representation preferred by the parser would be mapped onto the ultimate interpretation of the sentence. In other words, the structure endorsed by the parser should determine the interpretation of
the sentence. This traditional view is not unchallenged. For example, studies on *garden-path* sentences have revealed that comprehenders can obtain interpretations that conflict with the possible parses of the linguistic input (e.g. Christianson et al. 2001; Qian et al. 2018). The *good enough* approach to comprehension (Ferreira et al., 2001, 2002; Christianson et al. 2001; Ferreira & Patson, 2007) explains such findings by allowing interpretations derived through simple heuristics (e.g. world knowledge, word order, etc) rather than fully specified parses. The *noisy channel* account (Levy, 2008; Gibson et al. 2013), on the other hand, accounts for the empirical findings by introducing noise or uncertainty on the linguistic input a comprehender perceives.

The current study has two goals, one empirical and the other theoretical. First, we identify a new empirical case unrelated to the *garden-path* phenomenon, that demonstrates (descriptively speaking) misalignment between parsing and interpretation. Second, our account of the misalignment offers a new kind of analytical possibility to address the general question about the relationship between parsing and interpretation. Specifically, we will argue that interpretation should be modeled as the outcome of a comprehender’s pragmatic reasoning process. We follow the Rational Speech Act framework (RSA, Goodman & Frank, 2016) and apply bayesian pragmatic reasoning to account for our findings. The seeming conflict between parsing and interpretation will ultimately disappear, because in our analysis parsing preferences will be embedded under the pragmatic reasoning process to derive the ultimate interpretation.

Our case study concerns with the locality bias in sentence processing. In particular, we will examine the locality effect in wh-in-situ constructions that show scope ambiguity. More details about the wh-in-situ constructions will be introduced in the next section, but generally speaking, locality bias is commonly observed in sentence parsing. A representative example of this is the well-documented distance effect in processing non-local dependencies. In constructions that involve non-local dependencies, such as in English relative clauses or wh-questions, it is often observed that longer distance between the two elements on a dependency chain enhances processing difficulty, as measured by decreased acceptability judgments, increased reading time or enhanced neurophysiological responses (Gibson 1998; Warren and Gibson 2002; Van Dyke and Lewis 2003; Lewis and Vasishth 2005). As an
example, consider (1) from Alexopoulou and Keller (2007). In their results, as the distance between the verb *fire* and its fronted wh-argument *who* increased from (1a) to (1c), the acceptability rating decreased accordingly.

(1)  
   a. Who will we fire?  
   b. Who does Mary claim we will fire?  
   c. Who does Jane think Mary claims we will fire?

The shorter dependency is generally more preferred to the longer ones, hence the *locality* bias. Many accounts of this effect are based on hypotheses about how working memory is structured and deployed to support language comprehension. For example, in Dependency Locality Theory (Gibson 1998; 2000), the processing cost for completing a dependency is a function of the number of discourse references between the two elements on a dependency chain. Under the memory retrieval account (Lewis and Vasishth 2005), processing cost is in large part determined by how quickly and unambiguously the relevant dependent element can be retrieved from working memory, amongst all other memory representations that could potentially introduce interference. Longer dependencies are more likely to introduce elements that can interfere with the retrieval target, leading to an increased processing cost.

Although there is ample discussion in the literature about the parsing mechanisms for completing non-local dependencies, there is relatively little discussion about how the parsing outcome maps to the interpretation a comprehender obtains. It is more or less taken for granted that if a comprehender adopts a particular parse, she would also adopt the interpretation this parse generates. In this paper, using Mandarin wh-in-situ construction as our case study, we will show that the empirical picture is more complicated than this simple hypothesis would predict.

2. Parsing wh-in-situ scope – the locality bias

In Mandarin Chinese, a wh-in-situ language, a covert dependency is formed between an in-situ wh-phrase and its scope position (Aoun & Li, 1993; Cheng, 1991, 2003; Huang, 1982; Tsai, 1994). An example of a Mandarin Chinese wh-construction is given in (2):
The example in (2) presents an embedded wh-question: the wh-element *which official* takes scope on the left edge of the embedded clause. Despite the lack of overt cues that signal a non-local dependency, processing evidence from Xiang et al. (2015; 2020) showed that the incremental construction of a wh-in-situ dependency is constrained by the same parsing principles that regulate the processing of overt non-local dependencies.

More important for the current purpose, based on experimental evidence from eyetracking reading and acceptability ratings, Xiang & Wang (2020) argued that when there is scope ambiguity for a wh-in-situ element, the local scope dependency (low scope) is less costly than the non-local high scope dependency, essentially illustrating a locality bias like their overt-dependency kin in English. This conclusion is largely based on a comparison between two kinds of sentences, as shown by the examples in (3):

(3) a. 记者们  知道 [Clause1 市长  严惩了  哪些  官员。]  
    jizhemen zhidao  shizhang  yancheng-le naxie guanyuan  
    Reporter know  mayor  punish  which official  
    “The reporters knew which officials the mayor punished.” (High Scope) OR  
    “The reporters knew the mayor revealed which officials that the city council punished.” (Low Scope)  

b. 记者们  知道 [Clause1 市长  相信 [Clause2 市政府  严惩了  哪些-官员。]]  
    jizhemen zhidao  shizhang  xiangxin  shizhengfu  yancheng-le naxie-guanyuan  
    Reporter know  mayor  believe  city-council  punish  which-CL-official  
    “The reporters knew which officials the mayor untruthfully claimed that the city council punished.” (High Scope)  
    (unavailable): “The reporters knew the mayor untruthfully claimed which officials the city council punished.” (Low Scope, blocked)
The sentence in (3a) is ambiguous since the wh-in-situ item could take either high scope at the left edge of clause 1 or low scope at clause 2. The low scope, i.e. the local scope dependency that associates the wh-item with the lower clause 2, was argued by Xiang and colleagues to be more preferred to the high scope. The critical argument for this conclusion comes from the comparison between (3a) and (3b). The two sentences in (3a) and (3b) are almost identical, except that the lower verb believe in (3b) is lexically constrained such that it does not allow an embedded interrogative clause as its complement. Such a subcategorization constraint on verbs is well-known in the literature (Ginzburg, 1995), and we give some examples of verbs with distinct subcategorization properties in (4). Verbs like know or reveal allow either embedded interrogative or declarative complement clauses, as shown in (4a) and (4b). But verbs like believe or think only allow embedded declaratives, as shown by the contrast in (4c) and (4d).

    b. John knew/revealed Mary wrote that book.
    c. * John believed/thought who wrote that book.
    d. John believed/thought Mary wrote that book.

Given the verb difference between (3a) and (3b), one important consequence is that the lower scope dependency in (3b) is blocked. In Xiang and colleagues’ results, blocking the local scope dependency in (3b) led to substantial difficulty in processing, resulting in a much lower acceptability rating for (3b) than (3a) and longer regression reading time on the wh-morpheme in (3b) than (3a). The contrast between (3a) and (3b) strongly suggests a locality bias in identifying the scope position for wh-in-situ expressions. Importantly, additional follow-up experiments in Xiang & Wang (2020) also showed that the low acceptability for high-scope only sentences such as (3b) is indeed due to the unavailability of the low scope dependency, instead of the potential alternative account that (3a) may have simply benefited from the fact that it is scope ambiguous (e.g. the ambiguity advantage effect, Traxler, Pickering & Clifton, 1998). The critical observation is that if one switches the position of the verbs know and reveal in (3a), as well as the position of the verbs know and believe in (3b), the previously observed complexity differences between the two conditions would disappear.
The two new conditions are shown in (5).

(5)  

a. 记者们 透露了 [Clause1市长 知道 [Clause2市政府 严惩了 哪些-官员。]]

jizhemen toulu-le shizhang zhidao shizhengfu yancheng-le naxie-guanyuan
Reporter reveal-perf mayor know city-council punish which-CL-official

“The reporters revealed which officials the mayor knew that the city council punished.” (High Scope) OR

“The reporters revealed the mayor knew which officials that the city council punished.” (Low Scope)

b. 记者们 相信 [Clause1市长 知道 [Clause2市政府 严惩了 哪些-官员。]]

jizhemen xiangxin shizhang zhidao shizhengfu yancheng-le naxie-guanyuan
Reporter believe mayor know city-council punish which-CL-official

( unavailable): “The reporters believed which officials the mayor knew that the city council punished.” (High Scope, blocked)

“The reporters believed the mayor knew which officials the city council punished.” (Low Scope)

Parallel to (3a) and (3b), (5a) is scope ambiguous and (5b) is not. But the unambiguous (5b), critically different from the unambiguous (3b), only has the low scope parse, since the high scope is blocked by the matrix verb ‘believe’. There is no acceptability difference between (5a) and (5b), in contrast to the acceptability difference between (3a) and (3b).\(^3\) This contrast lends strong support to the conclusion that there is a locality bias in parsing. Whenever a local dependency is available, it is relatively easy for the parser to successfully establish a syntactic parse, as in the case of (3a), (5a) and (5b); but when the local dependency is blocked, as in (3b), the parser encounters a greater degree of parsing difficulty.

The ambiguity-based account, on the other hand, would make the wrong prediction that the ambiguous (5a) should be rated much higher than the unambiguous (5b).

The current study builds upon the observation that when there is scope ambiguity for

\(^3\)Based on the binary yes-no acceptability judgments reported in Xiang & Wang (2020, Experiment 2), sentences like (3a) and (3b) were rated on average at 0.7 and 0.3, and sentences like (5a) and (5b) were rated at 0.71 and 0.67.
wh-in-situ expressions, there is a strong preference for the local scope parse. The source of the locality bias is not the focus of the current paper. Xiang & Wang (2020) investigated the interaction between memory retrieval and syntactic expectation mechanisms to account for the locality bias. The primary goal of the current paper is to identify the interpretation bias people have for scope ambiguous wh-constructions, and explain how such interpretation bias arises. To start with, if interpretation bias tracks parsing bias, one reasonable hypothesis is that the scope ambiguity should ultimately be resolved to favor interpretations supported by the local scope dependency. We test this possibility in Experiment 1 using a truth value judgment task.

3. Experiment 1: Scope interpretation bias – A truth value judgment task

In this experiment, participants were presented with a sentence containing a wh-in-situ expression. The target sentence by itself has two possible interpretations: one is compatible with the low scope (local) dependency, and the other with the high scope dependency. But the participants were also presented with a context scenario that was only compatible with one of the interpretations. They were instructed to judge whether the target sentence fits the context. Their judgments, therefore, can provide us with some evidence as to which scope dependency they have committed to. Consider a target sentence like the one in (6):

(6) 艾米丽 公布了 [Clause1她的团队 发现了 [Clause2外星人 建造了 哪座城市。]]
Emily gongbu-le tade tuandui faxian-le waixingren jianzao-le na-zuo chengshi
Emily announce-perf her team discover aliens establish which-CL-city
High scope: “Emily announced which city her team discovered aliens established.”
OR
Low scope: “Emily announced her team discovered which city the aliens established.”

When the high scope reading is true, the sentence can be roughly paraphrased as “Emily announced the answer to the question ‘which city did Emily’s team discover the aliens established?’”. This reading entails that Emily revealed the identity of the city. Suppose the answer to the embedded question is “Rome”, then the high scope reading means that Emily revealed that her team discovered that the aliens established Rome. The low scope
reading, on the other hand, can be paraphrased as “Emily announced her team discovered the answer to the question ‘Which city did the aliens establish?’”. This reading, crucially, does not necessarily entail Emily revealed the identity of the city. This difference between the high and low scope dependencies will play an important role in our experiment below.

3.1. Material, participants and procedure

We constructed four different conditions. An example is shown in (7). These conditions share the same context scenario, and differ from each other in the target sentences. Participants were instructed to judge, after reading the context and the target sentence, whether the target sentence fits or does not fit the context scenario. For convenience, we will refer to the task as a truth value judgment task, and code the fit and the does not fit responses as true and false judgments respectively. The first condition (7a) has a target sentence like (7).

The preceding context makes the high-scope construal true and the low-scope construal false for the target sentence in (7a). To counterbalance the association between the True/False judgments and the high/low scope construals of the target sentences, we modified the matrix predicate in (7a) to create the condition in (7b). In (7b), the matrix verb is an antonym of the positive matrix predicate in (7a). We labeled the condition (7b) as “Matrix verb negative”, because for the majority of the stimuli items (12 items out of 16), the verb in condition (7b) contains an overt negation marker followed by a positive predicate. The context for (7b) is identical to (7a), but because the matrix verbs in these two conditions are antonyms, we expect the judgments provided to the target sentence should be the opposite. In this way we counterbalanced the the association between the True/False judgments and the high/low scope construals of the target sentences. In addition to the two ambiguous conditions in (7a) and (7b), we also included unambiguous target sentences as control comparison conditions, see (7c) and (7d). For these conditions, only the high scope reading is grammatically available, because the lower embedding verb, e.g. believe, blocks the low scope dependency. The unambiguous target sentences were preceded by the same context used for the ambiguous conditions, resulting in the judgment false for (7c) and true for (7d).

(7) Context: At a recent archaeology conference, Emily said that her research team found evidence to prove that a famous ancient city was actually built by aliens. But
she kept the name of the city a secret. (Chinese: 艾米丽说她的团队找到了证据证实某一个有名的古城市其实是外星人建造的。但目前她对这个城市的名字保密。)

Target sentence:

a. Ambiguous; Matrix verb positive

Emily announced which city her team discovered aliens established.

b. Ambiguous; Matrix verb negative

Emily hid which city her team discovered aliens established.

c. Unambiguous; Matrix verb positive

Emily announced which city her team believed aliens established.

d. Unambiguous; Matrix verb negative

Emily hid which city her team discovered aliens established.
Emily yinman-le tade tuandui xiangxin waixingren jianzao-le na-zuo chengshi
Emily hide her team believe aliens establish which-CL-city

High scope: “Emily hid which city her team believed aliens established.” (True)
Low scope: “Emily hid her team believed which city the aliens established.” (Unavailable)

We constructed a total of 16 sets of 4-condition items like (7a-d). The experiment was conducted on Ibex Farm (Drummond, 2016). For each trial, participants first viewed a context scenario, and then they pressed the space bar to view the target sentence on the next screen. On the target sentence screen, they could not go back to view the context scenario. They were instructed to decide, by choosing between two buttons presented to them on the screen, whether the target sentence fits or does not fit the given context. The 16 sets of experimental items were distributed to the participants in a Latin Square design, such that each participant only saw one condition from each item set. We also included 10 additional filler trials. The filler trials had the same format as the experimental trials, and 5 of them should be judged as true, while the other 5 as false. Ninety-eight native Mandarin speakers participated in our study, 10 of which were excluded because their response accuracy on the filler trials was lower than 60%. We report the results from the remaining 88 participants below.

3.2. Results

We first converted participants’ truth value judgments into whether they interpreted the target sentence with a high scope construal. For example, for (7a), a response of false was converted to high scope. The proportion of high scope choices is plotted in Fig. 1. There are more high scope responses for the unambiguous conditions (79% for the positive predicate condition and 77% for the negative predicate condition) than the ambiguous conditions (mixed effects logistic model: Est = −0.21 ± 0.08, z = −2.58, p < .01). This is unsurprising given that the unambiguous conditions can only be parsed as having a high scope for the wh-expressions. It is worth noting, however, that the proportions of high scope choices for

---

2We chose to use these particular wordings because the literal translation of true or false in Mandarin sounded unnatural as task instructions.
unambiguously high-scope sentences like (7c) and (7d) are not at ceiling. As we will show in Experiment 2, the unambiguous conditions tested here are syntactically complex and received very low acceptability ratings. The severe parsing difficulty on the unambiguous conditions may have led to inaccurate interpretation. But we note that the primary interest of the current study is to explain the interpretation bias for the ambiguous conditions, and the interpretation of the unambiguous conditions does not play a major role for the main purpose of the paper. For the current purpose, the more important finding from Experiment 1 is that the two ambiguous conditions both received overwhelmingly more high-scope responses, 73% for both the positive and the negative predicate conditions, significantly higher than the 50% chance level ($p < .0001$).

![Graph showing proportion of high scope choice](image)

Fig 1: Truth value judgment task results: proportion of participants’ choosing the high scope construal

### 3.3. Discussion of Experiment 1

Results from the truth value judgment task in Experiment 1 provided strong evidence that participants are predominantly biased towards interpreting an ambiguous wh-in-situ construction as having a high scope reading. This finding contradicts with any prediction based on a positive correlation between interpretation bias and parsing bias. As discussed in Section 1, there are good reasons to believe that from a parsing perspective, the local scope dependency (i.e. low scope) is less complex to establish and is the preferred parse for
the parser, and the non-local scope dependency (i.e. high scope) is more complex and less preferred. The interpretation bias revealed by Experiment 1, however, is the opposite of the parsing bias.

This conclusion, that the interpretation bias obtained in Experiment 1 is the opposite of the parsing bias, critically depends on the assumption that there is a locality bias in parsing, which is based on previous findings in Xiang et al. (2020). One potential concern is that although the constructions tested by Xiang and colleagues were the same as in the current study, the stimuli in the two studies are not exactly identical. In particular, a context scenario was included in the current study, but was absent in the previous study. The verbs used in these two studies are also not entirely identical. We therefore conducted an acceptability judgment experiment in Experiment 2 to replicate the parsing locality bias using the current set of stimuli.

4. Experiment 2: Reproducing the locality parsing bias in an acceptability rating task

4.1. Material, participants, procedure and predictions

The Experiment material for Experiment 2 was identical to Experiment 1, with a total of 16 sets of 4-condition experimental items (see an example in [7]) and 10 filler items. The experimental procedure was also identical to Experiment 1: each trial consisted of a context scenario followed by a target sentence. The only difference was that at the target sentence participants were instructed to make a binary judgment (Yes/No) as to whether the target sentence was acceptable or not. Thirty native Mandarin speakers participated in the study. We excluded 6 participants whose accuracy on filler trials was below 60%. The data analysis reported below was based on the remaining 24 participants.

If there is a parsing bias favoring the local scope dependency, we make the following prediction: For the ambiguous conditions (7a) and (7b), the local dependency is available, but for the unambiguous conditions (7c) and (7d), the local dependency is blocked. The locality bias for the lower scope could manifest in a higher acceptability for the ambiguous than the unambiguous conditions, since in the latter case the favored low-scope parse is
blocked and participants are forced to construct the disfavored high scope parse. It is well-known that parsing difficulty could significantly hamper acceptability ratings (e.g. Chomsky & Miller 1963; Hofmeister et al., 2013).

4.2. Results and discussion

The acceptability judgment results support our prediction that there is a local scope preference. As shown in Fig. 2, the unambiguous conditions were rated significantly less acceptable (mean 0.44) than the ambiguous conditions (mean 0.67) regardless of whether the predicate was positive or negative ($Est = -1.01 \pm 0.37, z = -2.78, p < .01$). Sentences with positive matrix predicates were also rated lower than those with negative matrix predicates ($Est = -0.82 \pm 0.34, z = -2.44, p < .05$). We did not have any hypotheses/predictions for the predicate difference, so we will not discuss the effects of the predicates any further; but we note that since the context scenario in general ends on a sentence describing what did not happen, e.g. Emily kept the name of the city a secret in (7), this may have primed participants to favor a negative predicate over a positive one in the target sentence for discourse coherence reasons.

![Fig. 2: Acceptability judgment results: Proportion of participants’ Yes responses](image)

\[\text{The converged mixed effects logistic model is: model=glmer(acceptability~ambiguity*verb polarity + (1+ambiguity|subj) + (1|item), data, family=binomial)}\]
The acceptability ratings for the ambiguous conditions (7a,b) reflect a moderate degree of complexity for these sentences, whereas the much lower ratings for the unambiguous conditions (7c, d) suggest severe parsing complexity. Both patterns replicate the previous acceptability rating results from Xiang & Wang (2020). The gradient acceptability in itself is not unusual for structurally complex sentences, given the well-established observations that even for completely grammatical sentences, heightened processing complexity can substantially reduce acceptability ratings (Chomsky & Miller 1963; Hofmeister et al., 2013). In the current case, the low ratings on the unambiguous conditions suggest that associating a wh-in-situ phrase with a non-local scope position is difficult and costly for comprehenders.

4.3. Summary of Experiment 1 and 2

The results from Experiment 1 and 2 present an empirical paradox. On the one hand, Experiment 2, using an acceptability rating task, confirmed the locality bias in parsing a wh-in-situ dependency, replicating similar findings from previous studies. In particular, sentences that make a local scope dependency available are judged much more acceptable than sentences that block the local scope dependency and allow only a non-local high scope dependency. Experiment 1 with a truth value judgment task, however, showed an anti-locality bias in the ultimate interpretation participants obtained for the same set of sentences. The apparent contrast between the results from these two experiments sets up the core empirical observation that parsing biases do not necessarily align with interpretive biases. It is important to note that, the claim about the misalignment between parsing and interpretation, i.e. a locality bias for parsing and an anti-locality bias for interpretation, is only relevant for the ambiguous conditions tested in this study. The unambiguous conditions mainly serve as

---

4The experiments in Xiang & Wang (2020) did not manipulate the polarity of the matrix verb, in fact, the majority of the items there have a positive matrix predicate. The basic sentence structures tested there are identical to the target sentences in (7), but they do not have a preceding context scenario. The mean acceptability judgments (yes/no binary judgments) reported in Experiment 1 of Xiang & Wang (2020) were 0.75 for the ambiguous condition, and 0.39 for the unambiguous condition; these ratings were replicated in their Experiment 2, with 0.7 for the ambiguous condition and 0.3 for the unambiguous condition. There were also completely ungrammatical filler sentences included in their study, which received an average acceptability rating of 0.18. The mean ratings for the ambiguous and unambiguous conditions from Xiang & Wang (2020) are on a par with the mean ratings in the current study, and the slightly lower ratings on the ambiguous positive predicate sentences in the current study are more likely due to the context scenario instead of the sentence complexity itself.
controls to help us detect the locality bias in parsing. The rest of the paper will therefore only focus on the ambiguous conditions and we will develop a formal proposal to reconcile the discrepancies observed in Experiment 1 and 2.

Language comprehension requires constructing structural representations for the linguistic input. This is the task of parsing. In the current case, a complete parse needs to specify where the scope position is for the wh-in-situ phrase. Needless to say, there is a close relationship between the parsing and interpretation, since semantic composition needs to be based on the parsed structures. But in the mean time, many other factors, among which pragmatic reasoning has been recognized as a salient one, could influence comprehension. The general idea that language communication should be viewed as a cooperative process between speakers and listeners, involving sophisticated pragmatic reasoning, is an old and extremely influential one (Grice, 1975). In recent years, this insight has been formalized using bayesian computational models, in particular the Rational Speech Act framework (RSA, Goodman and Frank 2016; Frank and Goodman, 2012). In the rest of the paper, we explore the possibility that the currently observed contrast between parsing and interpretative decisions can be (at least partly) captured by examining how a listener pragmatically reasons about the most likely messages the speaker has intended, given the form of the utterance, and the listener’s world knowledge. Our analysis makes use of the bayesian pragmatic reasoning model developed under the rational speech act framework. In section 5, we will first introduce some general background on this framework, and then extend the original model to the current case study. We will show that the basic formulation of bayesian pragmatic inferences could capture the qualitative patterns observed in the truth value judgment task. Built upon this result, we further develop an analysis in section 6 that demonstrates how parsing bias can be integrated into the pragmatic reasoning process, reconciling any apparent discrepancies between parsing and interpretation.

5. Deriving the truth value judgment results with bayesian pragmatic inferences

5.1. Pragmatic inferences in language communication

Linguistic utterances convey information about the world. A pragmatic listener, upon hearing an utterance, would update their probabilistic model of the world states based on the
information conveyed by the utterance. Following the recent rational speech act framework (Goodman and Frank, 2016; Frank and Goodman, 2012), we can model a pragmatic listener’s posterior belief about the world state \( w \) given the utterance \( u \), using the Bayesian inference, as shown in equation (8):

\[
P_L(w|u) = \frac{P_S(u|w) \times P_L(w)}{\sum_{w'} P_S(u|w') \times P_L(w')} \tag{8}
\]

The pragmatic listener (L) is conditioning their belief update on two factors. First, assuming the speaker S is cooperative and trying to be helpful, the listener works backwards and estimates the likelihood a speaker would have uttered \( u \) given the world state in the speaker’s mind (the term \( P_S(u|w) \)). Second, the listener also brings to the communication some prior belief as to how likely the world state \( w \) holds independent of the utterance (the term \( P_L(w) \)). The normalizing constant in (8) (i.e. the denominator), considers the alternative world states that the listener entertains as relevant when decoding the message delivered by the utterance \( u \). In this kind of framework, in order to model a pragmatic listener, we also need to model a pragmatic speaker, who chooses to utter \( u \) among a set of alternative utterances to describe a world state in her mind. The RSA model, therefore, captures the intuition that linguistic communication involves back-and-forth pragmatic reasoning between a listener and a speaker.

A speaker often has more than one choice when she linguistically encodes a message. In the RSA framework, a pragmatic speaker is assumed to be rational: she chooses her utterance from a set of alternative utterances according to the utility \( U_s \) that a particular utterance would obtain, as shown in (9). The pragmatic speaker would in general want to maximize her utility. The utility function could be defined in a number of ways (Goodman and Frank, 2016), and we follow the most basic definition that states a pragmatic speaker would choose to make the most informative utterance to the listener, as shown in (10). The equation in (10) measures how certain a listener is of a particular world state \( w \) upon hearing \( u \). Intuitively, if the speaker makes a very informative utterance, the listener should update her beliefs in such a way that the world state \( w \) intended by the speaker would become

\[^5\text{Strictly speaking, the utility function should also consider the cost of an utterance. For the sake of simplicity, we only consider the informativeness of an utterance here.}\]
more likely in the listener’s posterior beliefs. To avoid infinite recursion, the listener in (10) is defined to be a simple literal listener $L_0$, who updates her beliefs based on whether the literal meaning (i.e. the semantic meaning) of the utterance is true, as shown in (11).

$$P_S(u|w) \propto \exp(\alpha \times U_S(u; w))$$

$$U_S(u; w) = \ln(P_{L_0}(w|u))$$

$$P_{L_0}(w|u) = \frac{\delta_{[u]}(w) P(w)}{\sum_{w' \in W} \delta_{[u]}(w') P(w')}$$

The literal listener $L_0$ in (11) is crucial in order to connect the pragmatic reasoning process to the compositional semantics of the linguistic input. The term $\delta_{[u]}(w)$ in (11) takes the value 1 or 0, determined by whether the utterance $[u]$ is true or false when applied to a given world state $w$. All the world states that will make the utterance false will be removed, and the literal listener will update their beliefs based on the remaining world states (i.e. the ones that are compatible with the semantics of the utterance).

We will demonstrate in more detail in the following sections how each of the steps above can be implemented for the current empirical case. Our main interest is the ambiguous utterances $u$ in (7a) and (7b). These examples are repeated below in (12a) and (12b).

(a) Ambiguous; Matrix verb positive

Emily announce her team discover aliens establish which-CL-city

High scope: “Emily announced which city her team discovered aliens established.”

Low scope: “Emily announced her team discovered which city the aliens established.”

(b) Ambiguous; Matrix verb negative

Emily hide her team discover aliens establish which-CL-city

The free parameter $\alpha$ in (9) captures the extent to which the speaker is a rational agent, i.e. how much she would choose her utterance to maximize her utilities. There is no good way to theoretically determine the precise value of this parameter in advance. For the purpose of the current paper we do not need to set a particular value for $\alpha$. The speaker behavior will be experimentally estimated in section 5.4 and 5.5.
High scope: “Emily hid which city her team discovered aliens established.”
Low scope: “Emily hid her team discovered which city the aliens established.”

These utterances are ambiguous, and could convey information about different world states. If a listener’s comprehension process is modeled as updating her beliefs of each relevant world state \( w \) given an utterance \( u \), it is important to be clear what the relevant world states could be for the listener. The high or low-scope readings of the sentences above are semantic meanings derived from particular structural representations (i.e. depending on the scope dependency), and in principle, each of them could be mapped to one or more states in the world. Let’s first make clear what the most relevant world states could be for our working example in (12). When the matrix predicate is positive, as in (12a), the relevant world states are a set of possible combinations of two events: \( e_1 \): Emily announced the name of a city, which they discovered was built by aliens; and \( e_2 \): Emily announced their discovery that there was a city that was built by aliens. Let’s call \( e_1 \) the name announcement event, and \( e_2 \) the discovery announcement event. There are a total of 4 different ways to combine these two events, assuming each event takes either a true (+) or false (−) value, as shown in Table 1. Out of the 4 combinations, \( w_2 \) is not logically possible, since Emily couldn’t have announced the name of the city that they discovered was built by aliens without also announcing that they made such a discovery. In addition, \( w_4 \) is irrelevant since if neither event is true, the speaker wouldn’t have made the utterance in the first place. The two remaining world states \( w_1 \) and \( w_3 \) are therefore the two relevant states the pragmatic listener considers for the target sentence she hears. Applying the same reasoning to the target sentence with a negative matrix predicate, as in (12b), the relevant world states are also a set of possible combinations of two events: the name hiding \( e_1 \): Emily hid the name of a city, which they discovered was built by aliens; and the discovery hiding \( e_2 \): Emily hid their discovery that there was a city that was built by aliens. Among the 4 combinations of these two events, shown in Table 2, \( w_3 \) is logically impossible, because one can not hide the discovery of the city without also hiding the name of the city that was discovered. The possibility \( w_4 \) in Table 2 is again trivially irrelevant. The pragmatic listener would therefore consider two relevant world states \( w_1 \) and \( w_2 \) in Table 2 upon hearing the target sentence.

In Table 3, we summarize the remaining relevant world states considered by the listener.
Table 1: World states relevant for utterances with positive predicate

<table>
<thead>
<tr>
<th>world states</th>
<th>e1 name announcement</th>
<th>e2 discovery announcement</th>
<th>Considered as a relevant world state?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>$w_2$</td>
<td>+</td>
<td>−</td>
<td>no</td>
</tr>
<tr>
<td>$w_3$</td>
<td>−</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>$w_4$</td>
<td>−</td>
<td>−</td>
<td>no</td>
</tr>
</tbody>
</table>

Table 2: World states relevant for utterances with negative predicate

<table>
<thead>
<tr>
<th>world states</th>
<th>e1 name hiding</th>
<th>e2 discovery hiding</th>
<th>Considered as a relevant world state?</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>+</td>
<td>+</td>
<td>yes</td>
</tr>
<tr>
<td>$w_2$</td>
<td>+</td>
<td>−</td>
<td>yes</td>
</tr>
<tr>
<td>$w_3$</td>
<td>−</td>
<td>+</td>
<td>no</td>
</tr>
<tr>
<td>$w_4$</td>
<td>−</td>
<td>−</td>
<td>no</td>
</tr>
</tbody>
</table>

given the target sentences. The remaining relevant world states are relabeled in Table 3 as $w_1$ and $w_2$, and these are the $w_1$ and $w_2$ we will refer to in the later discussion. Note that for the positive and negative utterances, their corresponding $w_2$ states are essentially representing identical world affairs; but their corresponding $w_1$ states are different.

Table 3: A summary of the relevant world states considered in the model

<table>
<thead>
<tr>
<th>world states</th>
<th>Positive matrix predicate</th>
<th>Negative matrix predicate</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>Emily announced they discovered that a city was built by aliens and she also announced the name of the city. (艾米丽宣布了她们发现了有一个城市是外星人建造的，她也同时宣布了这个城市的名字。)</td>
<td>Emily hid the fact that they discovered that a city was built by aliens and she also hid the name of the city. (艾米丽隐瞒了她们发现了有一个城市是外星人建造的，她也同时隐瞒了这个城市的名字。)</td>
</tr>
<tr>
<td>$w_2$</td>
<td>Emily announced they discovered that a city was built by aliens but she did not announce the name of the city. (艾米丽宣布了她们发现了有一个城市是外星人建造的，但她没有宣布这个城市的名字。)</td>
<td>Emily did not hide the fact that they discovered that a city was built by aliens but she hid the name of the city. (艾米丽没有隐瞒她们发现了有一个城市是外星人建造的，但是她隐瞒了这个城市的名字。)</td>
</tr>
</tbody>
</table>
With the relevant world states defined as above, in this section we will model a pragmatic listener’s posterior belief about each world state given the utterance, that is, \( P_L(w_1|u) \) and \( P_L(w_2|u) \). Based on the equation in (8), in order to compute these posterior probabilities, one needs to know the prior probability for each world state, i.e. \( P_L(w_1) \) and \( P_L(w_2) \), and the likelihood for a speaker to produce the target utterance given the world state they have in mind, i.e. \( P_S(u|w_1) \) and \( P_S(u|w_2) \). We will first empirically estimate the prior probability of each world state in Experiment 3 (section 5.2). Then we will empirically estimate the production behavior in Experiment 4 (section 5.3). Finally in section 5.4 we derive the truth value judgments obtained in Experiment 1 using the empirically estimated priors and production data.

5.2. Experiment 3: Estimating the prior probabilities

5.2.1. Material, participants and procedure

To experimentally assess the prior probabilities of each different world state relevant to the listener, we first provided participants a neutral context that corresponds to the background scene used in Experiment 1, for example, a background scene about an archaeology conference. Once participants viewed the context sentence, on the next screen participants were instructed to choose between two possible situations that could take place in the given context. These two situations correspond to the two different world states illustrated in Table 3 (with different paraphrases). World states for sentences with positive and negative matrix predicates were tested in two different conditions in a within-subject design. The experiment material was closely modeled after material from Experiment 1. Sixteen sets of items corresponding to the original 16 sets of scenarios in Experiment 1 were constructed, with two conditions in each set of items. Each condition contains two choices (\( w_1 \) and \( w_2 \)). An example item set is given in (13) below. As we mentioned earlier in Table 3, the \( w_2 \) states under the positive and negative matrix predicates represent identical world affairs. We therefore used identical paraphrases for the \( w_2 \) situation under (13a) and (13b). In Figure 3 we also present an example trial.

(13) Context: At a recent archaeology conference, Emily made a presentation on behalf of her research team. (Chinese: 在最近的一次考古界的学术会议上, 艾米丽代表她
Question: Which of the following situation is more likely to arise? (Chinese: 以下的哪种情况更有可能发生？)

a. The positive predicate condition:

\( w_1 \): In her report, Emily said that her research team found evidence to prove that a famous ancient city was actually built by aliens. She also released the name of the city. (Chinese: 在她的报告里，艾米说她的团队找到了证据证实某一个有名的古城市其实是外星人建造的，她同时也宣布了这个城市的名字。)

\( w_2 \): In her report, Emily said that her research team found evidence to prove that a famous ancient city was actually built by aliens. But the name of the city needs to be kept secret for the moment. (Chinese: 在她的报告里，艾米说她的团队找到了证据证实某一个有名的古城市其实是外星人建造的，但目前她需要对这个城市的名字保密。)

b. The negative predicate condition:

\( w_1 \): Emily's research team actually has found evidence to prove that a famous ancient city was built by aliens. But in her report she didn’t mention this discovery at all. (Chinese: 艾米丽的团队其实已经找到了证据证实某一个有名的古城市是外星人建造的，但她在自己的报告里完全隐瞒了这个发现。)

\( w_2 \): In her report, Emily said that her research team found evidence to prove that a famous ancient city was actually built by aliens. But the name of the city needs to be kept secret for the moment. (Chinese: 在她的报告里，艾米说她的团队找到了证据证实某一个有名的古城市其实是外星人建造的，但目前她需要对这个城市的名字保密。)
At a recent archaeology conference, Emily made a presentation on behalf of her research team. (在最近的一次考古界的学术会议上, 艾米丽代表她的研究团队作了一个报告。)

(On a separate screen)

Which of the following situation is more likely to arise? (以下的哪种情况更有可能发生？)

1. In her report, Emily said that her research team found evidence to prove that a famous ancient city was actually built by aliens. She also released the name of the city. (在她的报告里，艾米丽说她的团队找到了证据证实某一个有名的古城市其实是外星人建造的，她同时也宣布了这个城市的名字。)

2. In her report, Emily said that her research team found evidence to prove that a famous ancient city was actually built by aliens. But the name of the city needs to be kept secret for the moment. (在她的报告里，艾米丽说她的团队找到了证据证实某一个有名的古城市其实是外星人建造的，但目前她需要对这个城市的名字保密。)

Fig 3: An example trial for Experiment 3. This example represents a trial that estimates the prior probability of each relevant world state under a positive predicate sentence.

The experiment was conducted on IbexFarm. A hundred and nineteen native Mandarin speakers participated in our study. The 16 sets of experimental items were distributed to participants with a Latin Square distribution, such that each participant only saw one of the two conditions for each item. There were also an additional 10 filler items, so each participant finished a total of 26 trials.

5.2.2. Results

Among the choices participants made for the positive predicates condition, there was on average a slight numerical preference for the $w_1$ state (0.53 $w_1$ vs. 0.47 $w_2$), not different from the chance performance ($p = 0.07$); for the negative predicates condition, there was a preference for $w_2$ over $w_1$ (0.42 $w_1$ vs. 0.58 $w_2$), significantly different from chance ($p < .0001$). With the prior probabilities of the relevant world states estimated, next, in Experiment 4 we estimate a speaker’s likelihood to produce the target wh-structure.
5.3. Experiment 4: Empirically estimating the production bias

5.3.1. Material, procedure and participants

The goal of this experiment is to estimate how likely participants will use the target wh-in-situ construction to describe a relevant world state. To this end, we first constructed scenarios that correspond to the four types of world states presented in Table 3. Next, we elicited productions that describe these world state scenarios. In particular, we are interested in whether participants will produce utterances identical or very similar to the ambiguous wh-in-situ target sentences used in the truth value judgment task in Experiment 1, as in (7a) and (7b). One methodological concern is that the target wh-in-situ construction is complex, and it is very unlikely that a free production task will trigger sufficient (or any) amount of target production. Previous production results from Xiang, Wang and Cui (2015) showed that native Mandarin speakers avoid producing relatively long wh-in-situ dependencies as much as they can, even at the cost of producing some otherwise dispreferred complex clause structures (e.g. relative clauses). Given this constraint, instead of eliciting free production, we provided phrase fragments to guide and constrain the participants’ production process.

We constructed a total of 16 item sets, with each item set containing 4 conditions, corresponding to the 4 relevant world states. These world states are very similar to the ones used in Experiment 3 (see an example in (13)). The experimental trials have the following structure. Participants saw one of the four world state scenarios on the screen. Below the context scenario on the same screen, they saw four phrase fragments. The participants were instructed to form a sentence using these fragments, which expresses a message coherent with the context scenario presented to them. The four fragments were presented in a 2x2 grid format, and the position of each fragment in the grid was randomized from trial to trial. For example, if a participant received a world state scenario for the positive predicate, e.g. either one of the two world states under (13a), the four fragments they would receive were “Emily announced”, “which city”, “established”, “her team discovered”. The same set of fragments were supplied to the participants for both the $w_1$ and the $w_2$ scenarios under the same positive predicate. If a participant received a relevant world state scenario for the negative predicate, e.g. either one of the two world states under (13b), they would receive an almost identical set of fragments except that the positive predicate “Emily announced” is
replaced by a negative one “Emily hid”. The positions of these fragments in the 2x2 grid were randomized so that participants were not cued about the word order of the target sentence they were about to produce. An example trial is given in figure 4. During the training phase participants were instructed that they could also add other material they want to use, as long as they include the provided phrases in their production. Even though the task itself is not equivalent to spontaneous natural production, it nevertheless leaves participants enough flexibility to form various types of utterances, and they were not overly forced to produce the target structure. The experiment material was adapted from Experiment 1 and Experiment 3. The world state scenarios were adapted from Experiment 3 (e.g. example (13)), and the phrase fragments were adapted from the target sentences in Experiment 1. The experiment was conducted on IbexFarm, and participants typed up and submitted each sentence they formed. A total of 248 native Mandarin speakers participated in our study. Each participant performed the task on 16 experimental trials and an additional 10 filler trials.

At a recent archaeology conference, Emily said that her research team found evidence to prove that a famous ancient city was actually built by aliens. She also released the name of the city. (在最近的一次考古界的学术会议上，艾米丽说她的团队找到了证据证实某一个有名的古老城市其实是外星人建造的。她同时还公开了这个城市的名字。)

Emily announced (艾米丽公布了)  which city (哪座城市)  
built (建造了)  her team discovered (她的团队发现了) 

Please make a sentence based on this scenario. The sentence you make should include the four phrases above, and its content should also be compatible with the scenario. (请根据这个场景造一个句子。您造的句子需要包括以上这四个词，还需要符合场景描述的内容).

Fig 4: An example trial for the production experiment reported in Experiment 4
5.3.2. Results

Three different native Mandarin speakers coded the production results. We removed the trials from participants that didn’t perform the task properly (about 1% of the total trials). For each trial, if the participant produced a wh-in-situ structure similar to the target sentence in the truth value judgment task in Experiment 1, it was coded as a target structure. Similarity was evaluated based on whether the four fragments provided to the participants were organized into the same word order and syntactic structure as the target sentences in Experiment 1. All other structures they produced were coded as non-target structures.

On average, about 40% of the total trials conformed to the wh-in-situ target structure, with similar word order as the target sentences used in Experiment 1. In Figure 5, we present the proportion of the target structure produced, split by the world state context and the predicate type. For both types of predicates, participants were more likely to produce the ambiguous target structure when describing the $w_1$ state than the $w_2$ state (positive predicate: $w_1$ 48%, $w_2$ 31%; negative predicate: $w_1$ 44%, $w_2$ 36%), as confirmed by a significant main effect of world state ($Est = 0.32 \pm 0.05, z = 6.8, p < .00001$).

---

The converged mixed effects model is: model=glmer(response~VerbPolarity*worldstate+(1|subj)+(1+worldstate|item),data=data, family=binomial). Both predictors are sum-coded.
It is worth noting that the target wh-in-situ dependency structure was not frequently produced by the participants (about 40% on average). This is not surprising given that the long wh-in-situ dependency is syntactically complex. Among the alternative structures participants produced, the most common strategy was to produce conjoined clauses such as “Emily announced her team discovered there was a city that was established by aliens, but she didn’t announce which city it was”. The conjoined-clause structure is longer in length than the target structure, but it is unambiguous. On some trials participants also produced structures that have the wh-phrase at the subject position, such as “Emily announced her team discovered which city was established by aliens”. This type of structure has shorter dependency length between the wh-phrase and the scope position, compared to the target structure; but it evokes a different kind of information structure packaging.

5.4. Deriving the pragmatic listener’s inferences

Finally, we are ready to work out the model prediction for the pragmatic listener’s posterior beliefs. When a pragmatic listener hears an ambiguous utterance containing a wh-in-situ expression, how would she update her beliefs about each relevant world states? We compute the listener’s belief update based on equation (8), repeated in (14) below.

\[
P_L(w|u) = \frac{P_S(u|w) \times P_L(w)}{\sum_{w'} P_S(u|w') \times P_L(w')}
\]

We have obtained information about the two terms \(P(w)\) and \(P_S(u|w)\) in Experiment 3 and Experiment 4. We will use those when calculating (14). For convenience, we first summarize the results from these two experiments in Table 4 and 5, with the positive and negative predicates separately, and then demonstrate how a pragmatic listener’s posterior probability is calculated.

---

8 The English translation may look like passivization, but the actual Mandarin production often involves a focus marker “shi”, instead of the passive marker “bei”, to front the wh-phrase to the clause initial position.
Table 4: For the positive predicate, see the example in (7a) and (12a):

Emily announced her team discovered aliens established which city.

(艾米丽公布了她的团队发现了外星人建造了哪座城市。)

| World states | $P_s(u|w)$ | Priors |
|--------------|-----------|--------|
| $w_1$: Emily announced they discovered that a city was built by aliens, and she also announced the name of the city. (艾米丽公布了她们发现了有一个城市是外星人建造的，她也同时公布了这个城市的名字。) | 0.48 | 0.53 |
| $w_2$: Emily announced they discovered that a city was built by aliens, but she did not announce the name of the city. (艾米丽公布了她们发现了有一个城市是外星人建造的，但她没有公布这个城市的名字。) | 0.31 | 0.47 |

\[
P_L(w_1|u_{\text{positive}}) = \frac{P_S(u|w_1) \times P_L(w_1)}{\sum_{w'} P_S(u|w') \times P_L(w')}
\]

\[
= \frac{0.48 \times 0.53}{0.48 \times 0.53 + 0.31 \times 0.47}
\]

\[
= 0.64
\]

\[
P_L(w_2|u_{\text{positive}}) = \frac{P_S(u|w_2) \times P_L(w_2)}{\sum_{w'} P_S(u|w') \times P_L(w')}
\]

\[
= \frac{0.31 \times 0.47}{0.48 \times 0.53 + 0.31 \times 0.47}
\]

\[
= 0.36
\]

Upon hearing the target utterance, the pragmatic listener’s posterior probability for $w_1$ is higher (0.64) than $w_2$ (0.36). Because the state $w_1$ is contradicting what the context scenario describes in Experiment 1 (see example (7a)), a higher posterior belief in $w_1$ predicts that participants would be more likely to answer False when they were asked in Experiment 1 whether the target sentence fits the given context. This correctly derives why participants appeared to show a preference for the high-scope compatible interpretation when they were presented with (7a).
For utterances containing a negative predicate, such as (7b), the calculation is very similar, as shown below.

**Table 5:** For the negative predicate, see the example in (7b) and (12b).

| World states | $P_s(u|w)$ | Priors |
|--------------|------------|--------|
| $w_1$: Emily hid the fact that they discovered that a city was built by aliens, and also hid the name of the city. (艾米丽隐瞒了她的团队发现了外星人建造的外星城，同时也隐瞒了这座城市的名称。) | 0.44        | 0.42   |
| $w_2$: Emily did not hide the fact that they discovered that a city was built by aliens, but she hid the name of the city.（艾米丽没有隐瞒她们发现了有一个城市是外星人建造的，但是她隐瞒了这个城市的名称。） | 0.36        | 0.58   |

(16)

$$P_L(w_1|u_{negative}) = \frac{P_S(u|w_1) \times P_L(w_1)}{\sum_{w'} P_S(u|w') \times P_L(w')}$$

$$= \frac{0.44 \times 0.42}{0.44 \times 0.42 + 0.36 \times 0.58}$$

$$= 0.47$$

$$P_L(w_2|u_{negative}) = \frac{P_S(u|w_2) \times P_L(w_2)}{\sum_{w'} P_S(u|w') \times P_L(w')}$$

$$= \frac{0.36 \times 0.58}{0.44 \times 0.42 + 0.36 \times 0.58}$$

$$= 0.53$$

Upon hearing the target utterance (7b), the pragmatic listener’s posterior probability for $w_1$ is lower (0.47) than $w_2$ (0.53). Because the preferred state $w_2$ is consistent with the context scenario in Experiment 1 (again see example [12b]), a higher posterior belief in $w_2$
predicts that participants would be more likely to answer *True* in Experiment 1. This again correctly derives the results that participants had a preference for the high-scope compatible interpretation when they were presented with (7b).

To summarize, in the discussion above, we assumed that a participant’s truth value judgment on a given trial is determined by her updated posterior beliefs of different world states. In particular, whether a participant gives a *True* or *False* response to a target sentence depends on which world state gains higher posterior probability, upon the participant’s encounter with the target sentence, and how the preferred world state fits the context scenario. Qualitatively speaking, the results from the calculations in (15) and (16) match the truth value judgment responses in Experiment 1. But a closer look still reveals some discrepancies. The proportion of the *True* or *False* responses from Experiment 1 is numerically larger than the posterior probabilities predicted for their corresponding world states: 73% empirical *False* responses vs. 64% predicted $w_1$ world state for the positive predicate condition in (7a); and 73% empirical *True* responses vs. 53% predicted $w_2$ world state for the negative predicate condition in (7b). Since the model prediction in (15) and (16) made use of empirically collected estimates for the production bias (Experiment 4) and the priors (Experiment 3), it is likely that there is a substantial amount of noise in our empirical estimates, which in turn affects the accuracy of the model prediction. But noise alone does not seem to explain everything. In particular, for those utterances containing a negative predicate, there is a large difference between the prediction and the empirical estimate, which seems to invite explanations beyond simply noise in the data. We do not have a fully developed account for this, but we will discuss a possible account based on QUDs in the General Discussion section.

### 6. Integrating parsing biases into the pragmatic reasoning process

In the last section, applying bayesian pragmatic reasoning, we derived the qualitative patterns of the observed truth value judgments. We have not shown, however, how the parsing bias, i.e. the parsing preference for the low scope instead of the high scope dependency, could be made consistent with the anti-locality bias reflected in the truth value judgments. In this section we develop an analysis to achieve this goal. The key of our analysis lies at
the level of the literal listener. As we introduced in section 5.1, a full bayesian pragmatic model carries out recursive reasoning between a listener and a speaker. A pragmatic speaker makes decisions about their production choices by reasoning about the linguistic update of a literal listener, and the outcome from the pragmatic speaker stage is in turn used to update a pragmatic listener’s beliefs at the next level. As the starting point of this chain of reasoning, the literal listener \( L_0 \) is the crucial step that connects structured semantic composition to pragmatic reasoning. In this section, we will first present an analysis that incorporates parsing biases into \( L_0 \)'s linguistic update. Then we will show how this analysis, with some limitations, makes correct qualitative predictions for the behavior of a pragmatic speaker.

6.1. Connecting parsing outcomes to pragmatic reasoning

A literal listener \( L_0 \) performs a belief update about different world states based on the literal meaning of a heard utterance. The basic formulation of \( L_0 \) in equation (11), adapted from the original RSA framework, only applies to utterances that are structurally simple and unambiguous. Extending it to deal with syntactically complex and ambiguous sentences, we assume the compositional semantics of an utterance \( u \) depends on how the surface string is parsed into different structures. In the current case, the target sentence has two possible parses, each representing one type of scope dependency. Let’s call the two parses \( u_h \) and \( u_l \), standing for the high-scope parse and the low-scope parse. We calculate \( L_0 \)'s inferences by combining different parses based on the weight/probability of each parse, as shown in (17):

\[
(17) \quad P_{L_0}(w|u) = P_{L_0}(w|u_h) \times P(u_h) + P_{L_0}(w|u_l) \times P(u_l)
\]

\[
= \frac{\delta_{[u_h]}(w)}{\sum_{w'} \delta_{[u_h]}(w')P(w')} \times P(u_h) + \frac{\delta_{[u_l]}(w)}{\sum_{w'} \delta_{[u_l]}(w')P(w')} \times P(u_l)
\]

Let’s consider our working example in (12a), in which the matrix predicate is positive.

---

9The derivation in (17) only holds because \( P(u) = P(u_h) + P(u_l) = 1 \), assuming that the current target utterance \( u \) only has two parses \( u_h \) and \( u_l \). The full derivation is the following: \( P_{L_0}(w|u) = \frac{P(w \cap u)}{P(u)} = P_{L_0}(w \cap u_h) + P_{L_0}(w \cap u_l) \) since \( u = u_h \cup u_l, u_h \cap u_l = \emptyset \), and \( P(u) = 1 \). \( P_{L_0}(w \cap u_h) = P_{L_0}(w|u_h) \times P(u_h) \), and \( P_{L_0}(w \cap u_l) = P_{L_0}(w|u_l) \times P(u_l) \).
The English glosses are repeated in (18). For convenience, we also repeat from Table 3 the two relevant world states assumed for this utterance.

(18) Emily announced her team discovered aliens established which city.

(艾米丽公布了她的团队发现了外星人建造了哪座城市。)

High scope parse: “Emily announced which city her team discovered aliens established.”

Low scope parse: “Emily announced her team discovered which city the aliens established.”

\( w_1 \) positive: Emily announced they discovered that a city was built by aliens and she also announced the name of the city.

\( w_2 \) positive: Emily announced they discovered that a city was built by aliens but she did not announce the name of the city.

Based on (17), we can compute the posterior probabilities a literal listener has for the world state \( w_1 \) and \( w_2 \) upon hearing the ambiguous utterance in (18). To start with, we will first make the simple assumption that it is equally likely for a literal listener to parse the ambiguous string in (18) into a high-scope or a low-scope dependency, i.e. \( P(u_h) \) and \( P(u_l) \) are equal at 0.5. We know this is in fact not true, since there is a locality bias in parsing that favors the low-scope parse (see Experiment 2 and the discussion there), and we will come back to modify this assumption in the end. If the utterance \( u \) is parsed as \( u_h \), it specifies the fact that the name of the city was made known. Under this parse the utterance is true with \( w_1 \) (hence \( \delta_{[u_h]}(w_1) = 1 \) in (19)) but false with \( w_2 \) (hence \( \delta_{[u_h]}(w_2) = 0 \)). If the utterance \( u \) is parsed as \( u_l \), since it underspecifies whether the name of the city is made known, it is compatible with both \( w_1 \) and \( w_2 \). We could not remove either \( w_1 \) or \( w_2 \) from consideration, and both are kept as viable options for the listener to consider. In addition, we already know the prior probabilities for \( P(w_1) \) and \( P(w_2) \) are 0.53 and 0.47 (see Experiment 3). The literal listener \( L_0 \) therefore updates her beliefs about \( w_1 \) and \( w_2 \) in the following way:

\[
(19) \quad \text{a. } P_{L_0}(w_1|u_{\text{positive}}) = \frac{\delta_{[u_h]}(w_1) P(w_1)}{\delta_{[u_h]}(w_1) P(w_1) + \delta_{[u_h]}(w_2) P(w_2)} \times P(u_h) + \frac{\delta_{[u_l]}(w_1) P(w_1)}{\delta_{[u_l]}(w_1) P(w_1) + \delta_{[u_l]}(w_2) P(w_2)} \times P(u_l)
\]
\[
\begin{align*}
&= \frac{1 \times 0.53}{1 \times 0.53 + 0 \times 0.47} \times 0.5 + \frac{1 \times 0.53}{1 \times 0.53 + 1 \times 0.47} \times 0.5 \\
&= 1 \times 0.5 + 0.53 \times 0.5 \\
&= 0.765
\end{align*}
\]

b. \( P_{L_0}(w_2|u_{positive}) \)

\[
\begin{align*}
&= \frac{\delta_{[u_h]}(w_2)P(w_2)}{\delta_{[u_h]}(w_1)P(w_1) + \delta_{[u_h]}(w_2)P(w_2)} \times P(u_h) + \frac{\delta_{[u_l]}(w_2)P(w_2)}{\delta_{[u_l]}(w_1)P(w_1) + \delta_{[u_l]}(w_2)P(w_2)} \times P(u_l) \\
&= \frac{0 \times 0.47}{1 \times 0.53 + 0 \times 0.47} \times 0.5 + \frac{1 \times 0.47}{1 \times 0.53 + 1 \times 0.47} \times 0.5 \\
&= 0 + 0.47 \times 0.5 \\
&= 0.235
\end{align*}
\]

This calculation suggests that even though the literal listener starts with a prior belief that the probabilities for \( w_1 \) and \( w_2 \) are very close to each other (0.53 and 0.47), after hearing the utterance in (18), the literal listener is leaning much more towards believing in \( w_1 \) over \( w_2 \).

When the utterance contains a negative matrix predicate, the working example from (12b) is repeated in (20). The calculation in (21) is very similar to the positive predicate case, but the compatibility between the utterance and each world state changes. When the utterance \( u \) is parsed as \( u_h \), it is compatible with both \( w_1 \) and \( w_2 \), hence both states need to be considered by the listener. If the utterance is parsed as \( u_l \), it is only compatible with \( w_1 \), and \( w_2 \) will be removed from further consideration. In addition, the prior probabilities for \( w_1 \) and \( w_2 \) were estimated to be 0.42 and 0.58 from Experiment 3.

(20) Emily hid her team discovered aliens established which city.

(艾米丽隐瞒了她的团队发现了外星人建造了哪座城市。)

High scope parse: “Emily hid which city her team discovered aliens established.”

Low scope parse: “Emily hid her team discovered which city the aliens established.”

\( w_1 \) negative: Emily hid the fact that they discovered that a city was built by aliens.
and also hid the name of the city.

Emily did not hide the fact that they discovered that a city was built by aliens, but she hid the name of the city.

(21) a. $P_{L_0}(w_1|u_{\text{negative}})$

$$= \frac{\delta_{[u_h]}(w_1)P(w_1)}{\delta_{[u_h]}(w_1)P(w_1) + \delta_{[u_l]}(w_2)P(w_2)} \times P(u_h) + \frac{\delta_{[u_l]}(w_1)P(w_1)}{\delta_{[u_l]}(w_1)P(w_1) + \delta_{[u_l]}(w_2)P(w_2)} \times P(u_l)$$

$$= \frac{1 \times 0.42}{1 \times 0.42 + 1 \times 0.58} \times 0.5 + \frac{1 \times 0.42}{1 \times 0.42 + 0 \times 0.58} \times 0.5$$

$$= 0.42 \times 0.5 + 1 \times 0.5$$

$$= 0.71$$

b. $P_{L_0}(w_2|u_{\text{negative}})$

$$= \frac{\delta_{[u_h]}(w_2)P(w_2)}{\delta_{[u_h]}(w_1)P(w_1) + \delta_{[u_l]}(w_2)P(w_2)} \times P(u_h) + \frac{\delta_{[u_l]}(w_2)P(w_2)}{\delta_{[u_l]}(w_1)P(w_1) + \delta_{[u_l]}(w_2)P(w_2)} \times P(u_l)$$

$$= \frac{1 \times 0.58}{1 \times 0.42 + 1 \times 0.58} \times 0.5 + \frac{0 \times 0.58}{1 \times 0.42 + 0 \times 0.58} \times 0.5$$

$$= 0.58 \times 0.5 + 0$$

$$= 0.29$$

The observation here is that even though the literal listener started with a lower prior probability for $w_1$ (0.42), after hearing the utterance, the listener’s posterior beliefs have changed to favor $w_1$ over $w_2$ (0.71 vs. 0.29).

The calculations in (19) and (21) showed that a literal listener, upon hearing a scope-ambiguous wh-sentence, would favor $w_1$ over $w_2$ regardless of whether the predicate is positive or negative. The calculations so far are based on the assumption that the literal listener has no parsing bias while parsing an ambiguous string $u$ into either a high-scope or a low-scope dependency ($p(u_h) = p(u_l) = 0.5$). This assumption needs refinement, since we already know the parser favors the low scope dependency over the high scope one. After we introduce the constraint $0 < p(u_h) < 0.5$ and $0.5 < p(u_l) < 1$ into the calculations in (19) and (21), it could be derived that for utterances with a positive predicate like (19), the literal listener’s posterior
probability for \( w_1 \) is between 0.53 and 0.765; and for utterances with a negative predicate like (21), it is between 0.71 and 1. In other words, upon hearing a scope-ambiguous target utterance, given the parsing preference that favors the low-scope dependency, the literal listener is predicted to assign higher posterior probability to \( w_1 \) than \( w_2 \), regardless of the polarity of the predicate.

6.2. From the literal listener to the pragmatic speaker

With the estimates for \( L_0 \)’s posterior probabilities, we can make a prediction as to how likely a pragmatic speaker would choose a target utterance \( u \) to describe a particular world state \( w \). As we discussed earlier (see (9)-(11)), the speaker’s choice of an utterance is largely determined by the informativity of this utterance, i.e., whether this utterance boosts the literal listener \( L_0 \)’s posterior belief of the target message/world state. As shown in (22), repeated from (9), the probability of a speaker choosing an utterance to deliver a message is proportional to the posterior probability that the \( L_0 \) listener would receive the target message from that utterance.

\[
P_S(u|w) \propto \exp(\alpha \times \ln(P_{L_0}(w|u)))
\]

Informally speaking, since \( L_0 \)’s posterior probability for \( w_1 \) is higher than \( w_2 \), we can make the prediction that a pragmatic speaker is more likely to use a scope ambiguous wh-in-situ target utterance, e.g. such as the ones in (18) and (20), to describe \( w_1 \) than \( w_2 \). This prediction is borne out in the empirically estimated production behavior in Experiment 4 (see Figure 5). The precise formulation of this prediction, however, is much less straightforward for a number of reasons. Below we present a simplified version of this prediction as well as the limitations of our simplified analysis.

The pragmatic speaker makes their production choices by comparing the utility of all the alternative utterances for a given world state. For simplicity, we have defined the utility in terms of the informativity of an utterance (see section 5.1). The contribution of the alternative utterances can be seen more clearly in (23), which is an extended version of (22).

\[
P_S(u|w) = \frac{\exp(\alpha \times U_S(u;w))}{\sum_{u' \in \text{ALT}} \exp(\alpha \times U_S(u';w))}
\]
Based on (23), given a well defined set of alternative utterances $u_1, u_2, ... u_n$ and a set of relevant world states $w_1, w_2, ... w_k$, one can calculate the production likelihood $P(u_i|w_j)$ for each pair of $u$ and $w$. The critical question is how to define the set of alternative utterances available to a speaker. Previous studies using the RSA framework often investigate syntactically simple structures, and it is relatively straightforward to define the set of alternative utterances for a speaker. For example, in the case of quantity implicature calculation that derives the *some but not all* inference from the quantifier *some*, it is reasonable to hypothesize that, *some* and *all* form the set of alternative expressions that a speaker could choose from.

In the current study, the target structure of interest is much more complex. It is difficult to define in advance, on a principled ground, the possible alternative structures a speaker may use. In fact, the empirical production data from Experiment 4 revealed a nuanced set of production strategies, which can also vary depending on the context scenarios and the lexical items involved. Another constraint, or limitation, of the current study is that we did not consider production cost when reasoning about the speaker’s production choices in (23). A speaker’s choice between alternative utterances more likely reflects a trade-off between the informativity of an utterance and the cost of that utterance. But even for simple utterances, there is no currently known satisfying metric for utterance cost. The problem is further complicated by complex syntactic structures like the ones investigated in this paper. As discussed in the result section of Experiment 4, on a substantial number of trials participants used unambiguous utterances that have a longer length than the target ambiguous form. There are also some trials for which participants used an ambiguous sentence with a different syntactic structure (information structure) from the target ambiguous form. It is unclear whether the utterance cost should be defined by sentence length, some metric about syntactic or information structure complexity, or the presence or absence of ambiguity.

Given these constraints, we make two simplified assumptions. First, we will only consider informativity but not cost when evaluating the utility of an utterance. Second, since a large number of the alternative structures produced by participants in Experiment 4 are unambiguous, we take the unambiguous utterances as the major alternative choices a speaker
has. We therefore assume three types of utterances available to the pragmatic speaker: the ambiguous target wh-in-situ construction $u_{\text{ambig}}$, the utterance $u_{\text{unambig}1}$ that unambiguously describes the world state $w_1$, and a different utterance $u_{\text{unambig}2}$ that unambiguously describes the world state $w_2$. The literal listener $L_0$’s linguistic updates based on these utterances are presented in Table 6.

### Table 6: Literal listener’s posterior probabilities for each pair of utterance and world state

<table>
<thead>
<tr>
<th>Alternative Utterances</th>
<th>$w_1$</th>
<th>$w_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$u_{\text{ambig}}$</td>
<td>$P_{L_0}(w_1</td>
<td>u_{\text{ambig}})$</td>
</tr>
<tr>
<td>$u_{\text{unambig}1}$</td>
<td>$P_{L_0}(w_1</td>
<td>u_{\text{unambig}1}) = 1$</td>
</tr>
<tr>
<td>$u_{\text{unambig}2}$</td>
<td>$P_{L_0}(w_1</td>
<td>u_{\text{unambig}2}) = 0$</td>
</tr>
</tbody>
</table>

Given the three types of alternative utterances in Table 6, which are an approximation but not a precise representation of all the possible utterances, based on (23) we derive the likelihood of a pragmatic speaker choosing the target wh-in-situ form $u_{\text{ambig}}$ to describe $w_1$ and $w_2$ in the following way:

\[
\text{PS}(u_{\text{ambig}}|w_1) = \frac{\exp(\alpha \times \ln(P_{L_0}(w_1|u_{\text{ambig}}))}{\sum_{u' \in \text{ALT}} \exp(\alpha \times \ln(P_{L_0}(w_1|u'))})
\]

\[
= \frac{\exp(\alpha \times \ln(P_{L_0}(w_1|u_{\text{ambig}}))}{\exp(\alpha \times \ln(P_{L_0}(w_1|u_{\text{ambig}})) + \exp(\alpha \times \ln(P_{L_0}(w_1|u_{\text{unambig}1})) + \exp(\alpha \times \ln(P_{L_0}(w_1|u_{\text{unambig}2}))}
\]

\[
= \frac{(P_{L_0}(w_1|u_{\text{ambig}}))^\alpha}{(P_{L_0}(w_1|u_{\text{ambig}}))^\alpha + 1 + 0}
\]

\[
= \frac{(P_{L_0}(w_1|u_{\text{ambig}}))^\alpha}{(P_{L_0}(w_1|u_{\text{ambig}}))^\alpha + 1}
\]

\[
\text{PS}(u_{\text{ambig}}|w_2) = \frac{\exp(\alpha \times \ln(P_{L_0}(w_2|u_{\text{ambig}}))}{\sum_{u' \in \text{ALT}} \exp(\alpha \times \ln(P_{L_0}(w_2|u'))})
\]

\[
= \frac{\exp(\alpha \times \ln(P_{L_0}(w_2|u_{\text{ambig}}))}{\exp(\alpha \times \ln(P_{L_0}(w_2|u_{\text{ambig}})) + \exp(\alpha \times \ln(P_{L_0}(w_2|u_{\text{unambig}1})) + \exp(\alpha \times \ln(P_{L_0}(w_2|u_{\text{unambig}2}))}
\]

\[
= \frac{(P_{L_0}(w_2|u_{\text{ambig}}))^\alpha}{(P_{L_0}(w_2|u_{\text{ambig}}))^\alpha + 1 + 0}
\]

\[
= \frac{(P_{L_0}(w_2|u_{\text{ambig}}))^\alpha}{(P_{L_0}(w_2|u_{\text{ambig}}))^\alpha + 1}
\]
Because we already know from the last section that $P_{L_0}(w_1|u_{\text{ambig}}) > P_{L_0}(w_2|u_{\text{ambig}})$, it can be derived that $P_S(u_{\text{ambig}}|w_1) > P_S(u_{\text{ambig}}|w_2)^{10}$. That is to say, the pragmatic speaker is more likely to use a scope ambiguous wh-in-situ target utterance when describing $w_1$ than when describing $w_2$. This is consistent with the empirical production results from Experiment 4.

In this section we developed an analysis that incorporates the parsing biases into the pragmatic reasoning process. In particular, parsing biases have an influence on the linguistic update of the literal listener, and this in turn makes correct qualitative predictions of the behavior at the pragmatic speaker stage. Due to a number of limitations of the current study, we only made qualitative but not quantitative predictions for the pragmatic speaker. But we hope our analysis nonetheless provides a principled way to integrate parsing and pragmatic reasoning under a unified framework.

7. General Discussion

There are two major findings in this paper. The first finding is an empirical one. Specifically, Experiment 1 and 2 identified an interesting paradox that was not previously observed. For a scope-ambiguous wh-in-situ construction in Mandarin, the parser prefers the local scope dependency, consistent with previous known parsing strategies recruited for dealing with many other types of long distance dependencies. The interpretive bias, however, points in the opposite direction: the interpretation compatible with the high scope dependency is the dominant interpretation. The pursuit of an explanation for this paradox led to the second main finding of this paper: A bayesian pragmatic model, following the rational speech act framework (Frank and Goodman, 2012; Goodman and Frank, 2016), could provide a principled (at least partial) explanation of the interpretation bias, while also incorporating

\[ \frac{(P_{L_0}(w_2|u_{\text{ambig}}))^\alpha}{(P_{L_0}(w_2|u_{\text{ambig}}))^\alpha + 0 + 1} = \frac{(P_{L_0}(w_2|u_{\text{ambig}}))^\alpha}{(P_{L_0}(w_2|u_{\text{ambig}}))^\alpha + 1} \]

---

10When $0 < y < x < 1$ and $\alpha > 0$, $x^\alpha y^\alpha + x^\alpha > x^\alpha y^\alpha + y^\alpha$, and it follows from there that $\frac{x^\alpha}{x^\alpha + 1} > \frac{y^\alpha}{y^\alpha + 1}$. 

37
the parsing bias into the model. In this section, we discuss the general implications of the current proposal and also some limitations.

7.1. Bridging the gap between parsing and interpretation

The experimental findings in the current study contribute new empirical evidence to the observation that there could be a misalignment between parsing and interpretation. Almost all existing approaches for addressing this issue focus on rethinking how parsing works. The good enough model (Christianson et al. 2001; Ferreira and Patson, 2007; ) proposes that the parsing outcome may not be a single complete parse, and interpretations can be derived through simple heuristics. The noisy channel account (Levy, 2008; Gibson et al. 2013) hypothesizes that there is uncertainty in the linguistic input a comprehender perceives, and some alternative input may become the candidate for parsing instead of the original input. The self-organizing model (Tabor, Galantucci, Richardson, 2004) allows a set of lexically anchored tree fragments to form a network via spreading activation, making it possible for locally coherent but globally ungrammatical parses to survive, which in turn explains why people sometimes accept interpretations that are not supported by the global parse. The approach we put forward in this paper departs from these previous approaches by rethinking instead the mapping between parsing outcome and interpretation. Our proposal is grounded in the idea that a comprehender’s task is not only to structurally represent the heard utterance, but also (or even more importantly) to infer a message or the communicative intent from the utterance. We maintain the basic parsing assumption that the original linguistic input is fully parsed into grammatical structures, but we introduce pragmatic reasoning to operate on the parsing outcome in order to derive the ultimate interpretation.11

We want to note that the current approach is not necessarily mutually exclusive with the other approaches mentioned above. It is also the case that the previous approaches were developed to account for a set of empirical observations different from the current empirical

---

11We also note that the proposal presented here, although incorporating the parsing outcome into the pragmatic reasoning process, does not aim to explain what leads to the parsing outcome in the first place. Following previous work, we assume there is a set of independent mechanisms that affect the parsing outcome, including the complexity of the to-be-established structure (Frazier and Fodor, 1978), working memory constraints (Gibson, 1998; Lewis, Vasishth and Van Dyke, 2006), syntactic or semantic expectation of the upcoming material (Hale, 2003; Levy, 2008), or contextual influence (Tanenhaus et al., 1995).
domain, for example, the classic garden path and local coherence phenomena. It is an open
question for future research whether the current approach could be extended to account for
these phenomena as well.

The current proposal adopts the rational speech act framework. The original RSA model,
focusing primarily on accounting for purely pragmatic phenomena such as scalar implica-
tures, deals with syntactically simple and unambiguous utterances (e.g. see a review in
Goodman & Frank 2016). In the current study, we extend the original model to syntacti-
cally complex domains. We assume a parallel parser that maintains multiple possible parses
of a sentence, with parsing biases represented as a probability distribution over all the possi-
ble parses. Each parse, according to its probability, has a certain level of activation strength.
But there is not a direct linear relationship between parsing biases and the interpretations
obtained by a listener. The linguistic update of a listener is determined by the interaction
between parsing biases and a number of other factors. To start with, in order to incorporate
parsing biases into the pragmatic reasoning process, we combine the effect of different parses
based on the probability of each parse when calculating the linguistic update at the literal
listener stage. As shown by the examples in (17), (19), and (21), three factors work together
to determine a literal listener’s linguistic update: parsing biases, the prior probabilities of the
relevant world states, and the compatibility between a world state and a particular parse of
the utterance. This means that parsing decisions alone do not necessarily determine a literal
listener’s linguistic update. Some parse is compatible with only one relevant world state, but
another parse may be compatible with more than one world state. Different world states also
have different prior probabilities. Due to the interaction between these different factors, even
when the parser strongly favors a particular parse, in principle it is still possible that the in-
terpretation (or the world state) supported by that parse does not become the dominant one
for the literal listener. Conversely, a world state compatible with a dispreferred parse still
has the chance to become a strong candidate in the posterior beliefs of a literal listener. This
is a key feature of the current proposal. It affords a more flexible mapping between parsing
and interpretation in a principled manner, allowing potential misalignment between parsing
preferences and interpretive preferences from the very beginning of the recursive pragmatic
reasoning process. The effect of the parsing bias, entering the pragmatic reasoning process
at the literal listener stage, will also eventually have an influence on the linguistic update of the pragmatic listener, mediated by the intermediate pragmatic speaker. But again the linguistic update of the pragmatic listener is also affected by other independent factors, in particular the prior belief held by the pragmatic listener for each relevant world state (see examples in (15) and (16)). Under this approach, parsing bias does not directly determine interpretation, instead, it becomes part of the overall pragmatic reasoning process that gives rise to the ultimate interpretation. The seeming paradox we observed earlier between parsing and interpretation in effect disappears.

It should be noted that the proposal we outlined here only integrates parsing and pragmatic reasoning at the global utterance level. It is an important open question how the general approach could be extended to incrementally integrate parsing and pragmatic reasoning, and thereby accounting for interpretations (of partial utterances) generated during incremental comprehension.

7.2. The potential role of QUDs

Although the bayesian pragmatic model provided good qualitative predictions for the interpretation bias, as we noted earlier, it did not completely capture the behavioral results from the truth value judgment task. The mismatch was more salient when the utterance contained a negative matrix predicate – the model predicted the true response for examples like (7b) to be only slightly above chance (53%) (see (16)), whereas the empirical result was substantially above chance (73%). This discrepancy suggests to us the current model needs further refinement. We speculate here that making the model more sensitive to the relevant questions under discussion (QUD, Ginzburg, 1996; Roberts, 1996) could potentially lead to improvement. A structured discourse can be perceived as being organized around a set of issues or questions that the interlocutors are committed to resolving together. Each sentence coheres with the previous discourse context by virtue of helping to address the currently shared (often implicit) QUD at that given moment in time, for instance, by providing an answer to it or by raising another relevant question. A comprehender would approach a given utterance as an answer to a discourse-salient QUD, and her pragmatic inference should be conditioned by this currently relevant QUD. A number of previous studies have explored how
to incorporate QUDs into the RSA models (Degen & Goodman, 2014; Savinelli, Scontras & Pearl, 2018; Scontras & Goodman, 2017). One empirical challenge with this approach is that QUDs are often assumed to be implicit, and there is no currently known rigorous method to systematically track QUDs in a discourse context. Even with this caveat, introducing QUDs into the current proposal could potentially still help better model the truth value judgment results. We sketch a few possible ways to do this below.

Recall that in our working example (7), the context scenario ended with a note that Emily kept secret the name of the city in their discovery. This last sentence may have made the naming event highly salient and relevant for at least some participants. These participants could be motivated to construct an implicit QUD like “Did Emily announce the name of the city?”. When they then received a target sentence and was asked to judge whether the target sentence fits the context scenario, they may have based their true/false judgments largely on how the target sentence answers this QUD and whether that answer is congruent with the context. In (15) and (16) we have computed the participants’ posterior beliefs about different world states after receiving a positive or negative target utterance. It is crucial to note that for an utterance containing a positive predicate, the two relevant world states in Table 4 would provide different answers to the QUD “Did Emily announce the name of the city?”. The $w_1$ state is a world state that will trigger the answer “Yes, she did” to the implicit QUD. This answer contradicts how the QUD was actually resolved in the context scenario, and therefore the comprehender would judge that the target sentence does not fit or false under the given context. The $w_2$ state, on the other hand, will trigger the answer “No, she didn’t” to the implicit QUD, consistent with how the QUD was resolved in the context scenario, leading to a truth value judgment fits or true. Since a pragmatic listener’s posterior belief was estimated to bias towards $w_1$, so far our prediction about the truth value judgment stays the same as before. But things change when we look at a target utterance containing a negative predicate. The two relevant world states in Table 5 would both trigger the same answer “No, she didn’t” to the implicit QUD, regardless of the listener’s posterior belief.

---

12A recent study from Ronai and Xiang (accepted) did an elicitation experiment to empirically identify potential QUDs pertaining to the calculation of scalar implicatures. In addition to uncovering questions that are consistent with the previous literature, their results also uncovered questions that have not been discussed as relevant to implicature derivation.
preferences for these two world states. This would mean that a participant should always conclude that the target sentence answered the QUD in a way consistent with how the QUD was resolved in the context. Therefore the target sentence has a very high probability to be judged as fits or true under the context. This could explain why in Experiment 1, the proportion of responding true for sentence containing a negative predicate is much higher than the model predicted posterior probability for the $w_2$ state in (16).

Under the scenario outlined above, the basic belief update process remains the same as our original proposal, and participants’ sensitivity to QUDs only has an effect at the last step of completing the truth value judgment task: instead of directly evaluating whether each updated world state is consistent with the context scenario, participants could evaluate whether each updated world state answers the discourse salient QUD in a way consistent with how the QUD is resolved in the context. Alternatively, it is also possible that QUDs could make contributions at a much earlier stage of the process. For instance, primed by the implicit QUD “Did Emily announce the name of the city?”, participants may decide to prioritize the parse that could clearly answer the QUD. Since only the high scope parse clearly specifies (the low scope parse underspecifies) whether the naming event happened, participants may be led to favor the high scope parse and give their truth value judgments based on the high scope parse. In this way, QUDs play a role in shaping participants’ early parsing decisions instead of only playing a role at a much later stage when participants make decisions about the truth value judgments. It is an attractive hypothesis that QUDs can have an effect on early parsing decisions, and this would be in keeping with some previous evidence (e.g. Clifton and Frazier, 2012). There is a potential challenge for this hypothesis, however, when the acceptability results from Experiment 2 are considered. Experiment 2 has identical material as Experiment 1, including the context scenarios. If contextually triggered implicit QUDs can guide participants to more readily settle on a high scope parse, this may incorrectly predict that participants could have overcome the locality bias in Experiment 2 and given higher acceptability ratings for sentences that only have a high scope parse (i.e. the unambiguous conditions).

Lastly, our discussion about QUDs so far still assumes an idealized listener who can build complete parses and integrate the parsing outcome with a pragmatic reasoning process.
There is yet another possibility. With complex sentences like the ones we tested here, participants may develop strategies to answer the implicit QUD “Did Emily announce the name of the city?” without fully parsing the target sentence. For example, they may have simply remembered the beginning of the target sentence “Emily announced” or “Emily hid”, and used those sentence fragments to answer the QUD, and to consequently derive the truth value judgments. We can not rule out this possibility. This possibility would provide a much simpler way to derive the preferences observed in the truth value judgment task. But the simplicity also comes with a theoretical disadvantage. Since this heuristic only narrowly targets the truth value judgment preferences, it is completely silent on how to account for the production preferences observed in Experiment 4, and an account of the latter has to be separately stipulated. The proposal we developed offers a more principled way to connect comprehension and production.

7.3. Analysis at the individual item level

The analysis we presented in section 5.4 was based on data aggregated over participants and items. It is worth asking whether the pragmatic model we used to explain the truth value judgments at the population level could also explain individual behavior. Unfortunately, as the truth value judgments, the prior estimates, and the production bias estimates in the current study were collected from different groups of participants, we are not able to construct a pragmatic model for each participant. But as a proof of concept, we nonetheless carried out a by-item analysis and examined whether the bayesian pragmatic reasoning introduced in section 5 could explain, at least to some extent, the truth value judgments obtained for each item.

Recall that in the current study we constructed 16 sets of scenarios/items like the ones presented in example (7). The same set of material, modified for the specific goals of different experiments, were used to collect truth value judgments, prior estimate and production bias estimates. We therefore could do the calculation in (15) and (16) for each item separately, and then correlate, at the individual item level, the posterior probability obtained for a world state and the truth value response consistent with that world state. Due to an experiment error, one item used to estimate the production bias in Experiment 4 had a slightly different
predicate from the same item used in the other experiments. We excluded this item from the by-item correlation analysis. The correlation results obtained from 15 items are plotted in Figure 6.

Fig. 6: By-item correlation between the truth value judgments from Experiment 1 and the posterior probabilities of the relevant world states.

In Figure 6, for each target sentence with a positive predicate (Figure 6, Left), we correlated the proportion of false (e.g. does not fit) responses with the posterior probabilities of the $w_1$ state. The y-axis in the figure represents the proportion of the false responses for each item. False is the majority response obtained for the positive predicates in Experiment 1. Since in Experiment 1 the $w_1$ state supports the false judgment for a positive predicate, the x-axis in the figure represents the posterior probability of the $w_1$ state. The calculation for the posterior probability is identical to the calculation of $P_L(w_1|u_{positive})$ in (15), except that it is now done for each individual item. Similarly, for each target sentence with a negative predicate (Figure 6, Right), we correlated the proportion of the true (e.g. fit) responses to the posterior probability of the $w_2$ state. True is the majority response obtained for the negative predicates in Experiment 1, and the $w_2$ state is the world state that supports the

Correlating the minority responses from Experiment 1 did not make a difference, e.g. correlating the true responses with the posterior probabilities of the $w_2$ states for the positive predicates.
true judgment. A significant correlation would indicate that, at the individual item level, the posterior probabilities of the relevant world states derived by the pragmatic model are indeed related to the experimentally estimated truth value judgments. But as shown in Figure 6, there are no significant correlations (ps>.3).

One possible source for the lack of correlation in the by-item analysis is the estimated prior probabilities for each world state. The scenarios we constructed for the current study are all somewhat arbitrary; at the individual scenario level, the prior probability estimate for each world state may have been too noisy to show a clear effect. We did an exploratory analysis that removed the effect of the prior from the calculation. This amounts to assuming an equal prior probability for the two alternative world states at the individual scenario level, with \( P(w_1) = P(w_2) = 0.5 \). The by-item correlation under this new analysis is presented in Figure 7. In Figure 7, for the positive predicate, the y-axis still represents the truth value judgments consistent with the world state \( w_1 \) (i.e. the proportion of false judgments); the x-axis, instead of representing the posterior probability of \( w_1 \), represents how likely a speaker would choose the target wh-in-situ structure to describe \( w_1 \), given that the speaker could use the target structure to describe either \( w_1 \) or \( w_2 \). For the negative predicate in Figure 7, the y-axis represents the truth value judgments consistent with the world state \( w_2 \) (i.e. the proportion of true judgments), and the x-axis represents how likely a speaker would choose the target wh-in-situ structure to describe \( w_2 \).

\[An additional methodological caveat, as pointed out by Whitney Tabor, is that Experiment 3, which estimated the priors, used neutral contexts. However, the truth value judgments in Experiment 1 were elicited under biased contexts. This also could have made the prior estimates less accurate for the purpose of modeling the truth value judgment results. Ideally, what we need is to measure participants’ a priori beliefs about different world states in the same context that they performed the truth value judgment task. In the current design, however, it is very difficult to overcome this problem because if the biased contexts from Experiment 1 were used for Experiment 3, they would immediately disambiguate between the choices presented to the participants.\]

\[One could see this more clearly based on the calculation of \( P_L(w_1|u_{\text{positive}}) \) in (15). When \( P(w_1) \) and \( P(w_2) \) are set to be 0.5 in this equation, the right hand side of the equation is essentially equivalent to \( \frac{P_S(u|w_1)}{P_S(u|w_1) + P_S(u|w_2)} \), and this is what the x-axis in Figure 7 (Left) represents. Similarly, for the plot on the right in Figure 7, the x-axis represents \( \frac{P_S(u|w_2)}{P_S(u|w_1) + P_S(u|w_2)} \).\]
The by-item correlation is marginal for the positive predicate items (p<.06), and significant for the negative predicate items (p<.05). This exploratory analysis provides some very preliminary evidence that at the individual item level, a listener’s truth value judgment is somewhat correlated with the production choice of the speaker. Overall, however, there is no strong conclusion we can draw at the individual item level. More future work is needed to address questions about individual variations.

8. Conclusion

To conclude, focusing on the wh-in-situ scope ambiguity in Mandarin Chinese, our study provides novel empirical evidence to show that parsing and interpretation decisions can misalign. We first demonstrate that by applying bayesian pragmatic reasoning to the data, we could derive the qualitative patterns of the truth value judgment results. We further develop an analysis that incorporates parsing decisions into a general pragmatic reasoning architecture, circumventing any actual conflict between parsing and interpretation. Our study therefore brings closer two strands of research in psycholinguistics, one on structure parsing, and the other on pragmatic reasoning.
Acknowledgment

The work reported here is supported by NSF BCS1451635.

References


