Word Order and Adjacency in the Processing of Nested Epistemic Expressions

Zhuang Qiu[ab] and Fernanda Ferreira[a] and Emily Morgan[a]

[a]University of California, Davis, United States; [b]Chinese University of Hong Kong, Hong Kong SAR

ARTICLE HISTORY
Compiled August 8, 2023

ABSTRACT
In this paper, we report three experiments investigating the cognitive mechanisms that underlie the semantic processing of a linguistic construction largely uninvestigated, namely, the nested structure of two epistemic modals in a single clause, as illustrated in the sentence "He *may certainly* have forgotten" (Lyons, 1977). Two theoretical approaches provide different account for the processing of this structure. Based on a formal linguistic account (Moss, 2015; Potsdam, 1998), the meaning of the second modal should be interpreted within the scope of the first modal, and thus, if the first and second modals switch their positions, a change in meaning would be expected. This account is referred to as the "scope account". In contrast, a good-enough processing account (Ferreira & Lowder, 2016) predicts that the scope of nested modals may not be thoroughly processed, and thus, the order of the modals should not change interlocutors' interpretation of the nested expression. We examined these two accounts in three experiments that elicited participants' interpretation of nested epistemic expressions, focusing on whether or not the order and the adjacency of the component modals affected how the nested expressions are interpreted. The repeated absence of the order effect was observed in all three experiments regardless of the adjacency of the modals. This finding challenges the scope account and suggests a holistic processing mechanism in line with the "good enough" processing framework.

KEYWORDS
Epistemic modals; scope; semantic processing; good-enough processing

1. Introduction

People think and behave not only based on their knowledge of reality, but also their expectation of what things might be like. The world view of uncertainty and probability forms an essential part of our everyday lives (Perkins, 1983, p. 6). The linguistic devices we use to communicate probability, words such as "certainly", "probable", and "might", are referred to as *epistemic modals* (Coates, 1983; Kratzer, 2012). As shown in Table 1, epistemic modals consist of words from different syntactic categories (parts of speech) such as adverbs, adjectives and auxiliaries. They serve as important means to modify the strength of an assertion (Hyland & Milton, 1997).

An important dimension in the meaning of epistemic modals is the strength of commitment to the factuality of the statement, which is termed "strength of modality"

CONTACT Zhuang Qiu. Email: zkqiu@ucdavis.edu

Word Categories	Epistemic Modals	Example Sentences
Adjectives	Certain, Probable, Possible	It is certain/probable/possible that he has forgotten the meeting.
Adverbs	Certainly, Probably, Possibly	He certainly/probably/possibly has forgotten the meeting.
Auxiliaries	Must, Would, Might	He must/would/might forget the meeting.

Figure 1. Epistemic modals in different word categories

by Huddleston and Pullum (2002, p. 175). For example, a speaker who says "He certainly has forgotten the meeting" shows a stronger commitment to his forgetting of the meeting than a speaker who says "He possibly has forgotten the meeting", and thus, the epistemic modal "certainly" has a higher (or greater) epistemic strength than the modal "possibly". This strength of commitment is influenced by the strength of evidence that supports the statement (Degen, Trotzke, Scontras, Wittenberg, & Goodman, 2019), and the spectrum of epistemic strength has been characterized by researchers using either an ordinal or a continuous scale. In some early studies, a three-point scale from high strength to medium strength and to low strength was adopted by Holmes (1982), Huddleston and Pullum (2002), and Halliday and Matthiessen (2004). Based on this classification, "certain (certainly)" and "must" express high degree of probability; "probable (probably)" and "will (would)" express medium probability, while "possible (possibly)" and "may (might)" express low probability (Halliday & Matthiessen, 2004, p. 116, 622). On the other hand, our current understanding of epistemic modality suggests that the strength of the modals should better be represented using a continuous scale from impossibility to certainty (Degen et al., 2019; Renooij & Witteman, 1999; Willems, Albers, & Smeets, 2019). The common ground shared by these two approaches is that the strength of different epistemic modals is graded and comparable. This study draws the strength from both approaches by adopting a numerical representation of epistemic strength in the rating experiments, while consulting the three-point representation when we categorized different modal sequences as "High" before "Low" or "Low" before "High", as shown in later sections.

An interesting phenomenon concerning epistemic modality that has not been widely studied is the use of two epistemic modals in a single clause, one being a modal auxiliary and the other being a modal adverb, as illustrated in the following examples (Lyons, 1977, p. 807):

1a. He *may possibly* have forgotten
1b. He *may certainly* have forgotten.

Following Moss's (2015) analysis of nested epistemic vocabulary, we name the usage of double epistemic modals like 1a and 1b as "nested epistemic expression". There is a noticeable difference between 1a and 1b in terms of the relative strength of the component modals. As Halliday (1970) noted, both of the epistemic modals in 1a have a low epistemic strength, and thus are considered as equivalent to each other. As to 1b, the first modal "may" expresses low probability while the second modal "certainly" expresses high probability, and in this case the strength of the two modals

is contradictory rather than equivalent to each other. That is why cases like 1a are called "modally harmonic" expressions while cases like 1b are called "modally non-harmonic" combinations (Lyons, 1977, p. 807).

Nested epistemic expressions often occur in casual communication. Qiu and Ferreira (2022) reported a corpus study in which researchers created a database of 400 thousand tweets and searched for nested epistemic expressions. They estimated that there is one case of nested usage out of a hundred cases of epistemic expressions. For modals used in casual conversations, it is important to draw a distinction between the nested epistemic expressions investigated in this research and the case of double modals like "might could", which often occurs in certain dialects of northern England and southern United States (Nagle, 2012). First of all, the component modals in a nested epistemic expression consist of an auxiliary and an adverb, while for the double modals, both of them are modal auxiliaries. Moreover, it is argued that the double modals contain one epistemic modal and one non-epistemic (such as deontic and dynamic) modal (Nagle, 1994), which is not the case for the nested epistemic expression. Last but not least, it has been argued that "might could" is one single lexical item consisting of two words, similar to a compound (Di Paolo, 1989), while nested epistemic expressions are compositional in that the two modals are believed to have different semantic scopes, and the second modal is believed to be embedded in the scope of the first modal (Moss, 2015; Potsdam, 1998).

Research on the processing of nested epistemic expressions sheds lights on our understanding of how the processor analyzes the scope of logical operations in daily communication. The notion of scope, which dates back to the Frege-Russell paradigm of semantics, is one of the most frequently used concepts in the study of language and logic (Hintikka, 1997). It describes "the relative priority of different logically active expressions" (Hintikka, 1997, p. 516). For example, "NOT EVERY student is happy" means something different from "EVERY student is NOT happy" because the scope of the negation "not" and the universal quantifier "every" differ in the above cases. In the first case, the universal quantifier is within the scope of the negation, and thus, the negation operates on the entire clause "every student is happy", meaning "it is not the case that every student is happy". In the second sentence, however, the negation is within the scope of the universal quantifier, meaning the quantifier operates on "student is not happy". In this case, the interpretation is "for every student, it is the case that the student is not happy". This example reveals that to derive the correct interpretation of the linguistic input, the processor needs to sort out the scope of different logical operations, which is often indicated by the word order.

Similarly, it has been argued that in a non-harmonic nested epistemic expression, one modal must be processed within the scope of the other modal (Lyons, 1977, p. 808). Moss (2015, p. 7) further explained that the strength of the outer (first) modal reflects the belief about the inner (second) modal. This is equivalent to saying that the first modal has a wider scope over the second modal in a nested construction. This idea echoed Potsdam (1998)'s syntactic analysis of the scope relation between nested modals, according to which, the first modal is higher in the syntactic structure than the second modal. In this paper, we refer to this account as "the scope account" of nested epistemic expressions. Based on this account, the meaning of "He certainly may have forgotten." differs from the meaning of "He may certainly have forgotten." To be more specific, the first sentence means "it is certainly the case that he may have forgotten". Notice that in this sentence, the statement "he has forgotten" is directly embedded within the scope of the modal "may", and then the whole expression "he may have forgotten" is further embedded within the scope of the modal "certainly".

By contrast, "He may certainly have forgotten" means "it may be the case that he certainly has forgotten". In this case, the statement "he has forgotten" is directly embedded within the scope of "certainly", and then the whole expression "he certainly has forgotten" is further embedded in the scope of the modal "may".

The implication of the scope account is that if we change the order of the component modals in a nested expression, the meaning of the entire expression will be changed. A possible aspect of meaning that could be changed by the order of the modals is the overall strength of the expression. The perceived probability of the statement "Tom has forgotten the meeting" may be different depending on whether that statement is embedded in "Tom may certainly have forgotten the meeting" or "Tom certainly may have forgotten the meeting". This hypothesis was tested in the first two experiments reported in this paper. The other possibility consistent with the scope account is that the meaning of the nested epistemic expressions depends on the order of the modals; however, the computation of the meaning is more complicated than the overall epistemic strength mentioned above. Moss (2015) argued that the first modal reflects the speaker's belief over the probability of the event described by the second modal. According to her argument, "Tom certainly may have forgotten the meeting" means "it is highly likely that Tom may have forgotten the meeting." This utterance can be truthfully made by a speaker who has observed many incidents of meetings that Tom should participate in, and for those meetings, Tom missed about half of them due to his forgetfulness. In this scenario, the modal "may" reflects the fact that it is possible for Tom to forget meetings, while the modal "certainly" reflects the speaker's confidence in Tom's forgetfulness given the total number of observations the speaker had. On the other hand, "Tom may certainly have forgotten the meeting" means "it may be the case that Tom has certainly forgotten the meeting." This utterance corresponds to a different scenario in which the objective probability for Tom forgetting the meeting is high (for example, Tom was extremely busy and the secretary failed to send a reminder), but the speaker's subjective estimate is uncertain, meaning that the speaker is not sure whether or not Tom has certainly forgotten the meeting. This interpretation of the scope account was tested in the third experiment of this paper.

The scope account claimed that the order of the nested modals determines how the entire expression is semantically interpreted. The question is, when interlocutors hear sentences like "Tom certainly may have forgotten the meeting" or "Tom may certainly have forgotten the meeting" in daily communication, do they pin down the scope difference implied by the word order and assign different meanings to the above cases? Or is it possible that interlocutors process the two modals holistically as an expression of uncertainty without attending to the order of the modals and the implied scope difference? Some classic studies on reading comprehension suggested that readers often do not faithfully interpret texts based on their face value. At the word level, Erickson and Mattson (1981) showed that an incorrect name in text often remains undetected if it semantically overlaps with the correct name. In their experiment, participants read questions like "How many animals of each kind did Moses take on the Ark?" and answered "two" without realizing that in the Bible, it is not Moses but Noah who took animals on the Ark. This kind of semantic illusion extends to the processing of noun phrase in general. Barton and Sanford (1993) found that when asking participants "When an airplane crashes, where should the survivors be buried?", the detection rate for the anomalous noun phrase was only about 30%. The shallow processing of linguistic input also occurs at the discourse level. Otero and Kintsch (1992) found that when reading a short paragraph containing contradictory information, participants tended to normalize the contradiction and interpret the text as a coherent piece. All

of the above suggested that comprehenders do not always achieve a veridical internal representation of the linguistic input they received (Traxler, 2014), which could also happen during the processing of nested epistemic expressions.

The good-enough theory of language processing (Christianson, Hollingworth, Halliwell, & Ferreira, 2001; Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Lowder, 2016) emphasizes the tendency that the parser processes information in a superficial manner, which sometimes leads to inaccurate interpretations. Syntax may be overridden in some cases, such as the misinterpretation of implausible passive sentences (like "the dog was bitten by the man") as active sentences (Ferreira, 2003; Gibson, Bergen, & Piantadosi, 2013); in other cases, the interpretation derived from a syntactic algorithm is influenced by the interpretation derived from heuristics, leading to a complex state of knowledge, a mixture of right and wrong interpretations (Ferreira & Lowder, 2016, p. 222). An example is the processing of garden-path sentences such as "While the man hunted the deer ran into the woods." Compared with a non-garden-path counterpart like "The deer ran into the woods while the man hunted", a garden path sentence is harder to process due to the reanalysis of initial misinterpretation. According to classic sentence processing theories (Frazier & Fodor, 1978; Frazier & Rayner, 1982), the parser initially treats the "the deer" as the object of the verb "hunted", and later reanalyzes it as the subject of the verb "ran". In this way, the correct interpretation of the sentence is that the deer ran into the woods, rather than the man hunted the deer. What Christianson et al. (2001, 2006); Ferreira et al. (2001) found was that although the parser eventually arrived at the correct interpretation, the mental representation of the revised parsing was not free from the previously experienced garden-path effect. When being asked whether or not the man hunted the deer after reading the garden path sentence, participants incorrectly answered "yes" about 40 percent of the time, which was significantly more frequent compared with the non-garden-path conditions. This is because the interpretation derived from the initial incorrect parse still lingers in memory without being completely replaced by the correct interpretation derived later in time (Huang & Ferreira, 2021; Slattery, Sturt, Christianson, Yoshida, & Ferreira, 2013).

Similar mechanisms may also take place during the processing of nested epistemic expressions. Instead of rigorously pinning down the scope of each component modal, the parser may perform superficial analyses by treating the occurrence of two epistemic modals as indicating uncertainty and hesitancy, while the order in which the modals occur does not change how the nested expression is interpreted. Qiu and Ferreira (2022) tested this hypothesis in three experiments in which participants read statements embedded in epistemic expressions and rated the probability of the statement based on the modals and the word order. Four experimental conditions were manipulated in which the experimental items either contained a single modal of high ("certainly has forgotten") or low ("may have forgotten") strength, or contained a nested expression in different word orders ("certainly may have forgotten" vs "may certainly have forgotten"). They found that statements received the highest probability rating were those embedded in the scope of a single "high" modal like "certainly"; while those embedded in the scope of a single "low" modal (like "may") received the lowest probability rating. Crucially, statements with nested epistemic expressions were rated in-between the probability of high modal and low modal conditions, and the order of the nested modals had no effect on the probability rating of the embedded statement.

The findings of Qiu and Ferreira (2022) do not support the scope account, and

are more consistent with the prediction of the good-enough processing framework. However, there are important questions to be answered before drawing the conclusion that the scope account does not explain the way people process nested epistemic expressions. First of all, Qiu and Ferreira (2022) attempted to study the effect of word order on probability rating with regard to the adjacency of the modals. In some of their experiments, the two nested modals were adjacent to each other (e.g. "He certainly may have forgotten"), while in another experiment, the two modals were separated by a parenthetical expression (e.g. "He certainly, according to the secretary, may have forgotten the important meeting"). It is hard to evaluate the effect of word order and the presence of a parenthetical (and also possible interactions) without including both factors in the same experimental design. Moreover, while the aim of their study was to explore the processing of nested epistemic expressions in casual communication, their experiments were based on the reading of written texts. Although the texts were written in the form of casual dialogues, participants were allowed to read the texts as many time as they wished to, and this paradigm does not closely mimic how nested epistemic expressions are processed in daily communication. Finally, Qiu and Ferreira (2022) focused on the overall strength of the nested epistemic expressions, asking whether expressions like "certainly may" and "may certainly" convey different probability. It is worth exploring other aspects of meaning beyond the overall epistemic strength to see whether or not they are contingent on the order of the modals.

This paper reported three experiments that addressed the limitation of Qiu and Ferreira (2022) and further investigated the processing of non-harmonic nested epistemic expressions. We asked the following questions which were designed to deepen our understanding of the cognitive mechanisms underlying the comprehension of this linguistic construction:

(1) First of all, in terms of the compositionality of the nested epistemic expression, what is the relation between the meaning of a nested epistemic expression and the meaning of the component epistemic modals, i.e. to what extent is the meaning of the nested expression derivable from the meaning of its individual components?
(2) To what extent do interlocutors process the scope of nested epistemic expressions during informal conversation? We examined the scope account and the good-enough processing account focusing on whether or not the order of the two modals has an effect on comprehenders' interpretation of the nested expressions. If participants interpret the nested expressions differently depending on the order of the modals, it means that the scope relation has been processed. However, if the order of the modals does not influence how participants interpret the expressions, then there is no evidence showing the scope is processed. If that is the case, the good-enough processing framework is a better description of the way in which nested epistemic expressions are processed.
(3) Last, does the distance of the component modals influence the way in which nested epistemic expressions are processed? Although nested epistemic expressions are semantically different from the double modal constructions, it is still possible that interlocutors interpret the component modals in a nested expression as a single lexical item if the two modals are adjacent, but not so when the two component modals are separated by other words. If that is the case, we should observe different interpretations of nested epistemic expressions depending on whether or not the component modals are adjacent.

2. Experiment 1

In this experiment, participants read English dialogues that contained epistemic expressions, and after reading each dialogue, they were instructed to rate the probability of a statement based on the information presented in the dialogue. In the dialogue, the order of the nested epistemic modals and the distance between them were manipulated to test whether readers interpret the probability of the embedded statement differently depending on the order and the distance of the modals. The scope account of the nested epistemic expressions would be supported by a difference in probability ratings between nested expressions with different word orders. Moreover, any difference should be larger when the distance between the two modals is expanded by a parenthetical phrase.

2.1. Participants

This study recruited 88 college students from psychology research participant pool at the University of California, Davis. All participants were native speakers of English, and naive concerning the purposes of the experiment.

2.2. Stimuli

This experiment contained 32 experimental items and 32 filler items, which were compiled into a questionnaire survey hosted by *Qualtrics* online survey platform. Each item in this experiment consisted of a written dialogue between two interlocutors, followed by a question asking participants to rate the probability of a statement based on the content of the dialogue. The format of the dialogue was the same across all 32 experimental items, in which, the first speaker asked a question, while the second speaker provided an answer to the question, and that answer contained an epistemic expression. As shown in Figure 2, the second speaker's reply fell into eight experimental conditions depending on the arrangement of the modals and whether or not a parenthetical element was included.

In terms of the arrangement of the modals, in the *High-Low* condition, the epistemic modal expressing higher probability (which is "definitely" in this example) preceded the modal expressing lower probability (which is "might" in this example). In the *Low-High* condition, the epistemic modal expressing lower probability preceded the modal expressing higher probability (such as "might definitely"). In addition to the nested-modal conditions, there were single-modal conditions in which only one epistemic modal was present. In the *High* condition, the sentence only contained the modal expressing higher probability (which is "definitely" in this example) and in the *Low* condition, the sentence contained only the modal expressing lower probability (which is "might" in this example).

It is important to note that the label of "high" or "low" is only relative to the pair of nested modals in question. For example, "probably" is the "high" modal when paired with "might" because the degree of probability expressed by "probably" is higher than that of "might"; however, the same modal "probably" becomes the "low" modal when paired with the modal "must" because the epistemic strength of "must" is supposed to be higher than that of "probably". The relative strength of the two modals was based on the theoretical account of Holmes (1982) and Halliday and Matthiessen (2004), and was confirmed empirically by this experiment. In the example shown in Figure 2, the

nested epistemic modals were "definitely might" (or "might definitely"), in which the modal expressing higher probability was "definitely", while the modal expressing lower probability was "might". Other experimental items may have a different combination of nested modals. In this study, we selected eight pairs of nested modals, which were "definitely and may", "definitely and might", "certainly and might", "certainly and may", "must and probably", "would and possibly", "probably and might", and "must and possibly". Each modal combination consisted of two epistemic modals, one expressing higher probability than the other (based on the survey of epistemic strengths in Qiu and Ferreira (2022), which is also confirmed in this study), and each modal combination appeared in four different experimental items. All of these modal combinations had been found in *Twitter* or *Corpus of Contemporary American English (COCA)*, indicating their validity as a feature of American English.

In addition to the arrangement of modals, experimental items also differed in terms of the presence of a parenthetical element indicating the source of information that the second speaker relied on when answering the first speaker's question. In this example, the parenthetical element was the prepositional phrase (PP) "according to the radio". Half of the experimental items had a parenthetical element, which was either placed in between the two modals in the nested-modal conditions or after the subject NP in the single modal conditions. For the other half of the stimuli, the parenthetical element was not included. Each experimental item ended with a question after the dialogue. The question probed participants' interpretation of what the second speaker said by asking the participant to judge the probability of a statement mentioned in the dialogue. In this example, the question asked about the probability of there being a traffic on Bob's way home. Since the statement "Bob has hit traffic on his way home" was embedded within the scope of different epistemic expressions in different conditions, participants' judgement would be different depending on which version of the stimuli they saw. After reading the dialogue, participants were instructed to indicate the probability using a slider from zero meaning "impossible" to 100 meaning "sure to happen".

Eight lists of experimental items were created following a Latin square design so that each list contained an equal number of items in each condition, while each experimental item only occurred once in a list. In total, there were 32 experimental items in each list, and each experimental item appeared in one of the eight conditions. In addition to the experimental items, each list also included 32 filler items to hide the intended research questions from the participants. Similar to the experimental items, each filler item was also a dialogue followed by a probability judgment question. However, the filler item did not contain any epistemic modals or it only contained the negation of an epistemic modal (such as "not likely"). Half of the filler items also contained a parenthetical element similar to that of the experimental items. During the experiment, all the items in the list were randomized. The inventory of all the items in this experiment can be accessed via GitHub[1].

2.3. Procedure

Participants used personal computers to access the web page hosting this online study, which started with demographic questions followed by the instruction and two practice trials. After the practice trials, participants hit a button to proceed. They were randomly assigned to one list of experimental items, and the first experimental item

[1] `https://github.com/PON2020/Nested_Epistemic_Expressions_Further_Inquiry`

"Bob hasn't arrived yet. Do you think he hit traffic on his way home?"		The arrangement of the modal(s)	The presence of parenthesis
Condition1	"Bob **definitely might** have hit traffic on his way home."	High-Low	No
Condition2	"Bob **might definitely** have hit traffic on his way home."	Low-High	No
Condition3	"Bob **definitely** hit traffic on his way home."	High	No
Condition 4	"Bob **might** have hit traffic on his way home."	Low	No
Condition 5	"Bob **definitely**, according to the radio, **might** have hit traffic on his way home."	High-Low	Yes
Condition6	"Bob **might**, according to the radio, **definitely** have hit traffic on his way home."	Low-High	Yes
Condition7	"Bob, according to the radio, **definitely** hit traffic on his way home."	High	Yes
Condition 8	"Bob, according to the radio, **might** have hit traffic on his way home."	Low	Yes
Question	How likely is it that there is traffic on Bob's way home?		

Figure 2. Experiment 1: Example of an experimental item

in the list was presented on the computer monitor. Participants read a dialogue and answered the question beneath the dialogue by moving a slider on the screen. After that, they hit the "proceed" button to reveal the next experimental item. Once the new experimental item revealed itself, there was no way to return to the previous item. The study ended after the participant completed all the judgment tasks in the list.

2.4. Data analysis

The design of this experiment was treated as 2x2, in which the first factor (the arrangement of modals) had four levels and the second factor (the presence of parenthetical elements) had two levels. The probability ratings of experimental items in different conditions were gathered and analyzed. Since the rating scores were bounded between zero and a hundred, we transformed the rating probability into its logit following the steps below:

(1) re-scale the rating score from *0-100* into *0-1*.
(2) re-code the re-scaled variable, for which one is coded as 0.999, while zero is coded as 0.001, following Verkuilen and Smithson (2012, p. 101).
(3) perform logit transformation on each re-coded rating score p using the equation:

$$logit(p) = log(\frac{p}{1-p})$$

Bayesian linear mixed-effects models were constructed using R package brms (Bürkner, 2017) with default priors [2] to explore the research questions listed in the Introduction section.

In the first model (Mod 1), we explored the compositionality of the nested epistemic expression by comparing the probability ratings of the nested conditions with those of the single modal conditions. This model treated the logit of probability ratings as a function of the modal arrangement (*High-Low, Low-High, Low, and High*) with a maximal random effects structures, including subject and item intercepts and slopes:

[2] The default priors for the slope of the fixed effect were flat priors, and more information on brms' default prior setting can be accessed following https://search.r-project.org/CRAN/refmans/brms/html/set_prior.html

$$(Mod\ 1)\ rating\_logit \sim 1 + modal\ arrangement + (1 + modal\ arrangement|subject) + (1 + modal\ arrangement|item)$$

The fixed effect was dummy-coded, and in order to compare each nested-modal condition with the single-modal conditions, we set the reference level of the predictor to the *High-Low* condition in one model (see Table 1), and to *Low-High* condition in another model (see Table 2), following the same model structure specified above. This model not only examined the strength of a nested epistemic expression with regard to the strength of its component modals, but also examined the effect of word order on how people interpret nested epistemic expressions in general.

In the second model (Mod 2), we focused solely on the data in the nested modal conditions and explored possible interaction effects between the word order variable and the variable concerning the adjacency of the component modals. We used word order (High-Low vs Low-High), the presence of the parenthetical element (present vs absent) and their interactions to predict the probability rating of nested epistemic expressions. The fixed effects were dummy coded, and maximal random effects structures were included. The data obtained from the experiments and the script used for statistical analysis is available in the GitHub[3].

$$(Mod\ 2)\ rating\_logit\_nested\_modals \sim 1 + word\ order * parenthetical + (1 + word\ order * parenthetical|subject) + (1 + word\ order * parenthetical|item)$$

2.5. Results

Participants' knowledge of the epistemic strength of each individual modal was reflected in the probability rating scores of the experimental items in single modal conditions. As shown in Figure 3, participants' understanding of the degree of certainty expressed by various epistemic modals revealed a gradient scale. When the epistemic modals "may", "possibly" and "might" appeared in the dialogue, participants rated the statement as reflecting the least certainty (about 65 out of 100). On the other hand, when the word "definitely", "would" and "certainly" were used in the dialogue, participants rated the statement as reflecting the highest certainty (more than 85 out of 100). The probability rating of model "must" (78.3) and "probably" (70.6) lay in between the previous two groups. The rating scores of all the epistemic modals in this study were above 60 out of 100, showing that the participants interpreted the strengths of these epistemic modals as above the chance level.

The mean probability rating scores for statements in different experimental conditions are shown in Figure 4. In this study, the statements in the *High* condition were rated above 80 out of 100, while the statements in the *Low* condition were rated below 70. The rating scores of the nested-modal conditions lay in between the rating scores of single-modal conditions, with *High-Low* condition rated slightly lower than the *Low-High* condition. Two Bayesian linear mixed-effects models were constructed to predict probability rating scores as a function of modal arrangement, following the

[3] https://github.com/PON2020/Nested_Epistemic_Expressions_Further_Inquiry

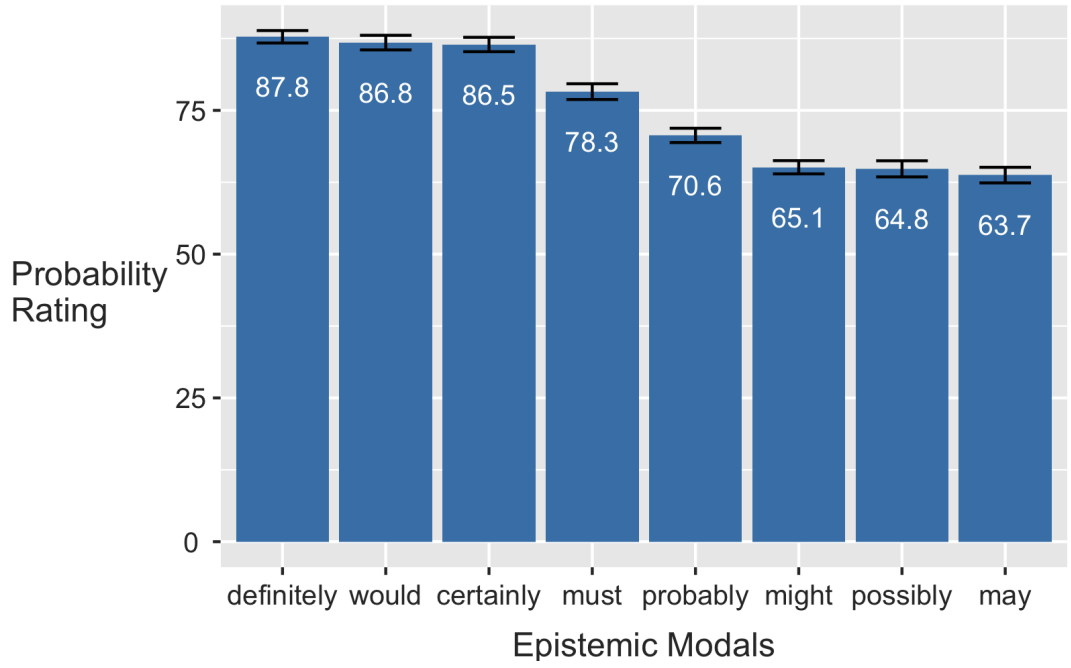


Figure 3. Experiment 1: Mean probability ratings of different epistemic modals (with standard errors). The ratings were averaged across the single modal conditions that contained the modal in question regardless of whether or not a parenthetical element was presented.

steps listed in Section 2.4. The structures of these two models were almost identical (see Section 2.4 Mod 1) with the only difference being the baseline for comparisons: *High-Low* condition was the baseline in one model (see Table 1) while *Low-High* condition was the baseline in the other model (see Table 2). We found that items containing a single modal of higher epistemic strength received the highest probability rating (β = 1.21, 95% CI = [0.82, 1.61] as in Table 1, and β = 1.17, 95% CI = [0.78, 1.55] as in Table 2), while items containing a single modal of lower epistemic strength received the lowest probability rating (β = -0.34, 95% CI = [-0.50, -0.18], and β = -0.39, 95% CI = [-0.55, -0.23]). This pattern was consistently observed regardless of the baselines used for comparison.

A closer look at the two nested-modal conditions revealed that the small difference between the *High-Low* and *Low-High* condition in the probability rating was not statistically meaningful. When *High-Low* condition was the baseline for comparison (see Table 1), it was rated lower than the *High* condition (β = 1.21, 95% CI = [0.82, 1.61]) but higher than the *Low* condition (β = -0.34, 95% CI = [-0.50, -0.18]). Importantly, the 95% credible interval for the slope of *Low-High* contained the value of zero (β = 0.05, 95% CI = [-0.10, 0.20]), meaning that there was no reliable statistical difference between the *High-Low* condition and *Low-High* condition. The same pattern was observed when the *Low-High* condition was made the baseline for comparison (see Table 2). The probability rating for both the *High* condition (β = 1.17, 95% CI = [0.78, 1.55]) and *Low* condition (β = -0.39, 95% CI = [-0.55, -0.23]) was statistically different from the baseline, but not the *High-Low* condition (β = -0.05, 95% CI = [-0.21, 0.11]).

To study whether or not the adjacency of the component modals influences the processing of nested epistemic expressions, we constructed a Bayesian linear mixed-effects

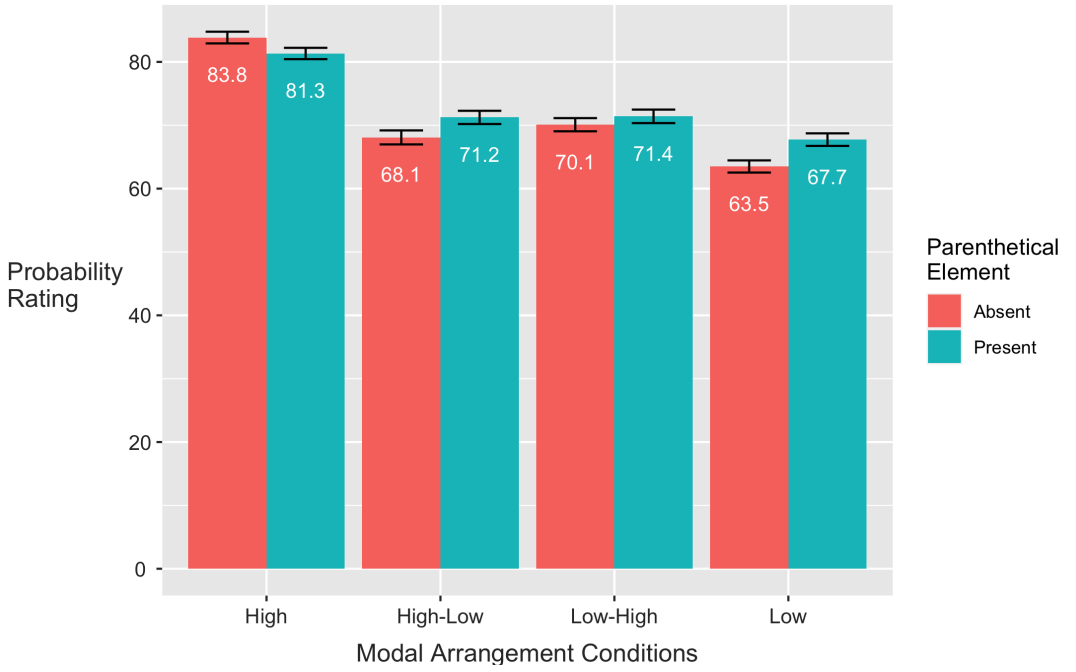


Figure 4. Experiment 1: Mean probability rating of different experimental conditions (with standard errors)

Table 1. Experiment 1: The main effect of modal arrangement estimated using a Bayesian linear mixed-effects model with High-Low condition as the baseline

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.18	0.14	0.90	1.46
Low-High	0.05	0.08	-0.10	0.20
High	1.21	0.20	0.82	1.61
Low	-0.34	0.08	-0.50	-0.18

[a]An estimate is statistically meaningful when zero is not included within the 95% credible interval.

Table 2. Experiment 1: The main effect of modal arrangement estimated using a Bayesian linear mixed-effects model with Low-High condition as the baseline

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.23	0.14	0.97	1.49
High-Low	-0.05	0.08	-0.21	0.11
High	1.17	0.20	0.78	1.55
Low	-0.39	0.08	-0.55	-0.23

[a]An estimate is statistically meaningful when zero is not included within the 95% credible interval.

model using word order (High-Low vs Low-High), the presence of the parenthetical element (present vs absent) and their interactions to predict the probability rating scores (see Mod 2 in Section 2.4). The output of the model is summarized in Table 3. We can see from this table that nested epistemic expressions with *Low-High* word order were rated slightly higher than the baseline *High-Low* order, but the difference was not statistically meaningful ($\beta = 0.17$, 95% CI = [-0.04, 0.38]). Adding a parenthetical element to enlarge the distance between the two modals increased the probability rating of the nested expressions in the *High-Low* condition ($\beta = 0.24$, 95% CI = [0.03, 0.45]), however, the interaction effect between word order and the adjacency of the

component modals was not statistically meaningful (β = -0.24, 95% CI = [-0.51, 0.03]). This means the order of the modals did not change the rating of the nested epistemic expression regardless of whether or not a parenthetical element was presented.

Table 3. Experiment 1: The output of Mod 2, the effect of word order, modal adjacency, and their interaction on how people process nested epistemic expressions

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.06	0.16	0.75	1.36
Low-High	0.17	0.11	-0.04	0.38
With-Parenthetical	0.24	0.11	0.03	0.45
Low-High:With-Parenthetical	-0.24	0.14	-0.51	0.03

[a]An estimate is statistically meaningful when zero is not included within the 95% credible interval. The baseline for comparison is the High-Low condition without a parenthetical element.

2.6. Discussion

When the participants read a statement embedded in the scope of an epistemic modal, the probability rating of that statement reflected participants' knowledge about the strength of that modal. The findings of this experiment showed that participants were sensitive to the epistemic strength of the modals in the dialogue, and thus they made reasonable inferences about the probability of the embedded statement. For instance, the sentence "Tom certainly has forgotten." and "Tom may have forgotten." both indicate the probability of *Tom has forgotten*. This means the same proposition "Tom has forgotten" is embedded within the scope of the epistemic modal in both of the sentences. The difference between these two sentences lies in the semantics of the modal. While "certainly" expresses high probability, the word "may" expresses low probability. Participants were able to calculate the probability of the statement by applying the semantics of the epistemic modal to the statement within its scope, and thus, in this experiment, the statements embedded in the *High* condition were rated noticeably higher than the same statements in the *Low* condition (see Figure 4 in Section 2.5, and also Figure 11 in the Appendix).

The epistemic strength of each modal, as reflected from the probability rating, revealed a gradient scale from high probability to intermediate probability and then to low probability. An interesting difference was found between participants' rating and the theoretical perspectives of Halliday and Matthiessen (2004) with regard to the strength of the modal "would" and "must". While Halliday and Matthiessen (2004, p. 116, 623) proposed that the word "would" expressed medium probability and "must" expressed high probability, the reversed pattern was observed among the participants in this study. In general, participants interpreted "would *S*" (the letter *S* stands for an embedded statement) as an expression similar to "it is definitely the case that *S*" or "it is certainly the case that *S*", rather than "it is probably the case that *S*"; on the other hand, they rated "must *S*" as an expression similar to "it is probably the case that *S*" which has a lower strength than what Halliday and Matthiessen (2004) believed. It is important to note that Halliday and Matthiessen (2004) was published twenty years ago and was not based on American English, and thus it is not unexpected that some epistemic modals are interpreted slightly differently from what was believed twenty years ago. Given that the focus of this study is on the processing of nested epistemic expressions, namely the *High-Low* and *Low-High* conditions, our experimental manipulation makes sense as long as the modals within each pair of nested combination express different degree of probability. That is the case of our

experiment because, as mentioned, the probability rating of the *High* modal condition was indeed higher than that of the *Low* modal condition.

As to the nested epistemic expressions, the scope account predicts that if the order of the two component modals has changed, the overall meaning of the nested expression will change accordingly. Thus, a difference in probability rating between the *High-Low* and *Low-High* conditions would support the scope account. In this experiment, we found that statements embedded in the nested-modal conditions were rated higher than the statement embedded in the single *Low* modal condition but lower than statement embedded in *High* modal condition. This pattern was observed regardless of whether the *High-Low* (see Table 1) or *Low-High* condition (see Table 2) was the baseline for comparison. This suggested that the epistemic strength of the nested-expression lies in between the epistemic strength of the component modals. However, there was no difference between *High-Low* and *Low-High* conditions in probability rating. For nested epistemic expressions, changing the word order of the two modals did not change the way people interpreted the probability of the embedded statement. Thus, we did not find evidence for the scope account in this experiment.

The lack of ordering effect held regardless of the presence of the parenthetical elements. This means sentences like "Bob certainly, the neighbour suggested, may be in the candy shop" receives the same probability rating as "Bob may, the neighbour suggested, certainly be in the candy shop". The absence of interaction effect between the word order and the presence of parenthetical element addressed the possibility that the scope of the epistemic expression would be processed when the two nested modals were not adjacent to each other. This possibility needs consideration because one type of double-modal constructions in some southern American dialects, such as "might could", has been argued to be a single lexical item (Di Paolo, 1989). It follows that the nested-epistemic expressions investigated in this study might also be treated as a single lexical item unless the two modals were separated. It turned out that the adjacency of the component modals did not influence the way people process nested epistemic expressions.

Although there wasn't an interaction effect between the parenthetical element and word order in the nested-modal data, the single-modal data suggested an interaction effect between the strength of the individual modal and the presence of parenthetical elements. When a statement was embedded within a single modal of lower epistemic strength, adding a parenthetical element increased the probability of the statement; by contrast, when a statement was embedded within a single modal of higher epistemic strength, adding a parenthetical element decreased the probability of the statement.

The parenthetical elements used in this experiment were expressions indicating the source of information upon which the second speaker's statement was based, such as "based on what I heard", and "according to the forecast", etc. These parenthetical elements belong to the notional category of evidentials, which are expressions or grammatical markers that "indicate something about the source of the information in the proposition" (Bybee, 1985, p. 184). In general, adding the source of information changes the credibility of the message as a whole. For example, compare "Bob, according to the radio, might have hit traffic on his way home" with "Bob might have hit traffic on his way home". If we believe that the radio is a reliable information source, adding the parenthetical evidential should make the epistemic expression "Bob might have hit traffic on his way home" more credible. What we found in this experiment was that for stimuli in the single *Low* modal condition, the perceived probability of the embedded statement increased when the information source was added to the epistemic expression, while for stimuli in the *High* modal condition, the perceived

probability of the embedded statement decreased when the information source was added to the epistemic expression. We argued that the evidential expression functions as a hedge, which softens the strength of strong statements, while boosts the strength of weak statements. A possible computational mechanism underlying this process will be illustrated in the General Discussion session.

The aim of this experiment was to examine the mechanism of processing nested epistemic expressions during casual conversations. Since the nested epistemic expression is less frequent in formal registers, we designed the experimental stimuli in a way that mimicked informal conversations, in which the nested epistemic expressions are more frequent and sound more natural. An important difference between the experimental set-up and natural conversation is that in this study the experimental stimuli were presented visually, and there was no time limit in the probability judgment task, making it possible for participants to read the stimuli for multiple times before proving their answers. If that happened, the findings of this experiment may not represent the cognitive processes that take place during casual conversations. Thus, it is necessary to replicate this experiment in a modified paradigm that more closely resembles informal daily conversations.

3. Experiment 2

This experiment attempted to replicate the first experiment using a modified paradigm that resembled daily informal conversations. We created a more natural context to investigate the processing of nested epistemic expressions focusing on the effect of word order and the adjacency of the modals.

3.1. Participants

109 college students were recruited from psychology research participation system at the University of California, Davis. They participated in the study in exchange for course credits. All participants were native speakers of English and naive concerning the purposes of the experiment. Participants were excluded if their accuracy rate for attention-checking items was lower than 90%. As a result, the data from 80 out of 109 participants were included in the data analyses.

3.2. Stimuli

All the stimuli in the previous experiment described in Section 2.2 were read and recorded by two native speakers of American English. For each dialogue, the line of the first interlocutor was read by the female speaker, while the line of the second interlocutor was read by the male speaker. The recordings were edited using the software program Praat (Boersma, 2011), which standardized the intensity of all the stimuli to 70 dB. We also spliced the modals and the parenthetical elements into the stimuli to control for their acoustic properties. The question following the dialogue was not included in the recording, and was presented visually.

The filler items in this study were also adopted and recorded from those in Experiment 1. We selected 13 filler items as the attention-checking items to detect possible disengagement from the tasks during the experiment. Those checking items were also dialogues followed by probability judgment questions, while the probability

of the statements was either extremely possible or extremely impossible, and thus, if a participant's rating was in the opposite direction, that trail would be regarded as incorrect. An example of the checking items is as below:

Female speaker: "I feel very hungry. Do you think we can be seated in an hour?"
Male speaker: "We are No.1 on the waitlist, and according to the waitress, they are preparing a table for us now."
Question: How likely is it that the speakers will be seated in an hour?

We can tell based on the reply of the male speaker that the interlocutors will be seated very soon, and it is almost certain that they will not wait for more than an hour. Thus, the probability rating of this item is supposed to be greater than 50 out of 100. If the rating score given by a participant was less than 50 for this checking item, we would mark it as incorrect. If a participant's accuracy rate for the attention-checking items was below 90%, the data of that participant would be excluded from the analyses.

3.3. Procedure

Participants used personal computers to access the web page hosting this online study, which started with demographic questions followed by the instruction and three practice trials. The practice trials mimicked the format of the experimental items in which the dialogues were presented acoustically and the questions were displayed visually. Participants were instructed to press the play button to hear the recorded dialogue, and after that, they read the question and rated the probability of a statement in the way described in Section 2.3. The recordings in this experiment could be played only once. After the practice trials, participants hit a button to proceed. They were randomly assigned to one list of experimental items, and the first experimental item in the list was presented on the computer monitor screen. Participants listened to the dialogue and provided a probability rating the same way as they did in the practice trials. After that, they hit the proceed button to reveal the next experimental item. Once the new item revealed itself, there was no way to return to the previous item. The study ended after the participant completed all the judgment tasks in the list.

3.4. Data analysis

This experiment followed the same data analysis procedure as the first experiment (see Section 2.4).

3.5. Results

The probability ratings of individual modals are shown in Figure 5. Similar to the first experiment, we see a gradient scale of epistemic strength with the word "definitely" and "certainly" at the high end of the scale and the word "might" and "may" at the low end. On the other hand, a few differences in the rating scores of some specific epistemic modals are evident. Compared with the first experiment, the rating of the modal "would" decreased from 86.8 out of 100 to 80.4 out of 100, moving away from the high end of the scale to the intermediate position of the scale. At the low end of the scale, the rating of modal "may" and "might" decreased slightly from about 64 out of 100 to about 60 out of 100. Although there were small changes in the rating score of

some modals, it is important to note that the relative strength difference between the two modals in a nested expression was preserved, meaning that the *High* modal in a nested expression was still rated higher than the *Low* modal in the nested expression (see Figure 6 below, and also Figure 12 in the Appendix).

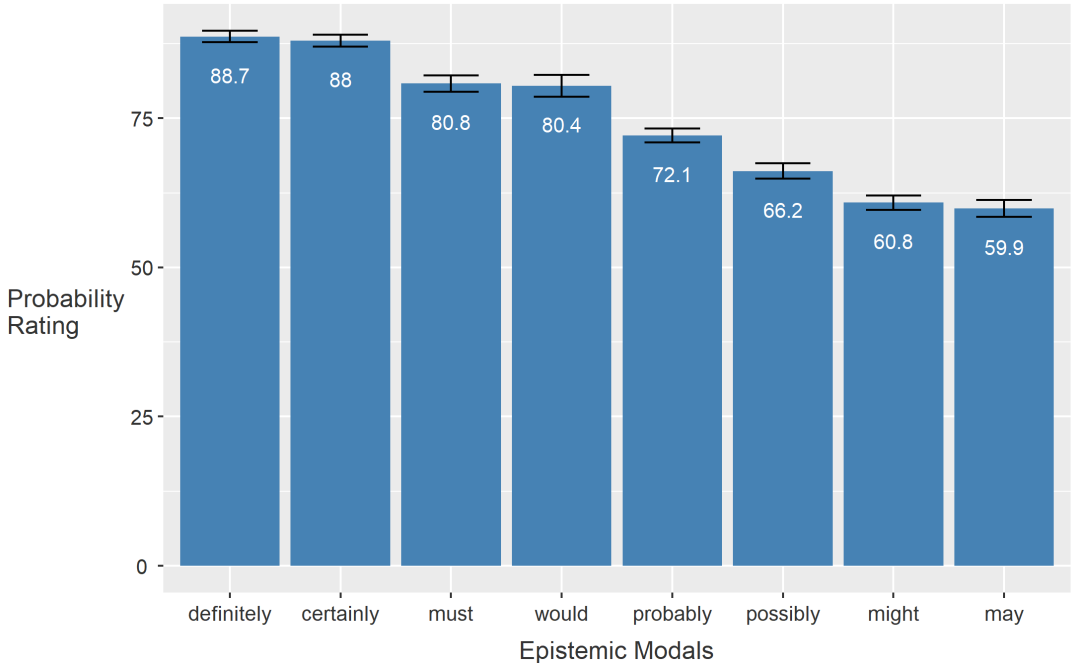


Figure 5. Experiment 2: Mean probability ratings of different epistemic modals (with standard errors). The ratings were averaged across the single modal conditions that contained the modal in question regardless of whether or not a parenthetical element was presented.

The probability rating scores in different modal-arrangement conditions echoed the patterns we found in Experiment 1. In general, statements embedded within the scope of a single *High* modal received the highest rating (83.8 and 82.7); while the statements embedded within the scope of a single *Low* modal received the lowest rating (60.3 and 66.7). Statements in the nested-modal conditions were rated in-between the ratings of the *High* and *Low* conditions, with the *Low-High* condition being rated slightly higher than the *High-Low* condition. Bayesian linear mixed-effects models were constructed estimating the logit of the probability rating score across conditions following the steps illustrated in Section 2.4. With regard to the effect of modal arrangement, when the *High-Low* condition was the baseline for comparison (see Table 4), it was rated lower than the *High* condition (β = 1.60, 95% CI = [1.25, 1.95]), but similar to the *Low* condition (β = -0.10, 95% CI = [-0.25, 0.05]). When the *Low-High* condition was the baseline for comparison (see Table 5), it was rated lower than the *High* condition (β = 1.47, 95% CI = [1.10, 1.85]) but higher than the *Low* condition (β = -0.22, 95% CI = [-0.36, -0.08]).

A closer look at the nested-modal conditions showed that the order of the two modals did not influence the rating of the embedded statement. When *High-Low* condition was the baseline for comparison, the 95% credible interval for the slope of *Low-High* condition included the value of zero (β = 0.12, 95% CI = [-0.02, 0.25]), meaning that there was no reliable statistical difference between the *High-Low* condition and *Low-High* condition. When the *Low-High* condition was the baseline, the 95% credible

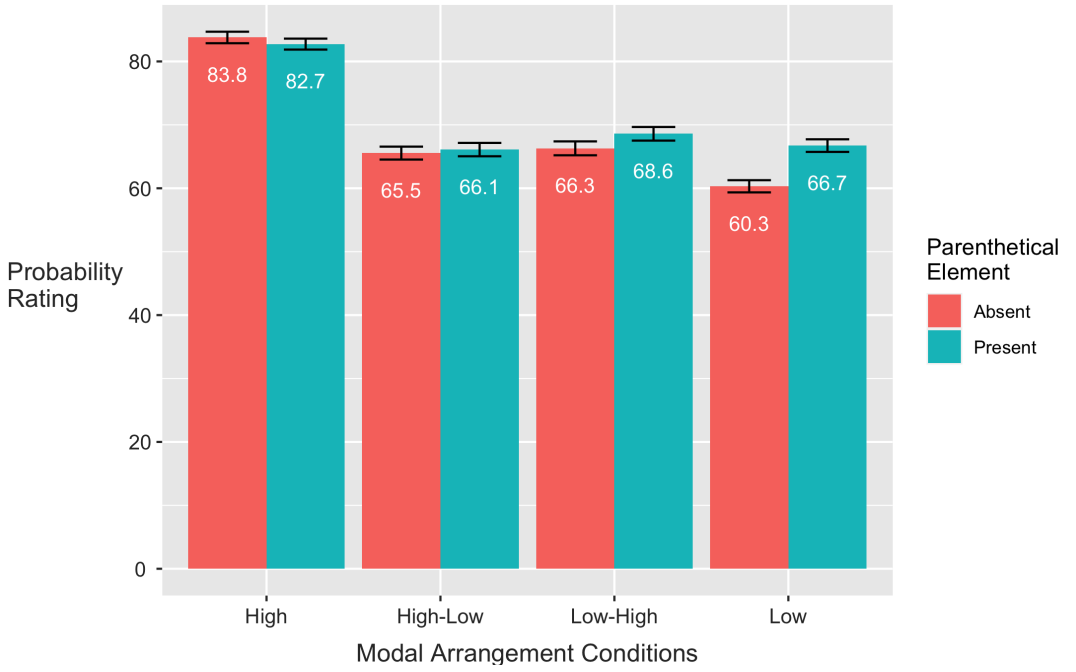


Figure 6. Experiment 2: Mean probability rating of different experimental conditions (with standard error)

Table 4. Experiment 2: The main effect of modal arrangement estimated using a Bayesian linear mixed-effects model with High-Low condition as the baseline

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.84	0.12	0.61	1.07
Low-High	0.12	0.07	-0.02	0.25
High	1.60	0.18	1.25	1.95
Low	-0.10	0.08	-0.25	0.05

[a]An estimate is statistically meaningful when zero is not included within the 95% credible interval.

Table 5. Experiment 2: The main effect of modal arrangement estimated using a Bayesian linear mixed-effects model with Low-High condition as the baseline

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.96	0.12	0.73	1.18
High-Low	-0.12	0.07	-0.26	0.02
High	1.47	0.19	1.10	1.85
Low	-0.22	0.07	-0.36	-0.08

[a]An estimate is statistically meaningful when zero is not included within the 95% credible interval.

interval for the slope of *High-Low* also included zero (β = -0.12, 95% CI = [-0.26, 0.02]).

We constructed a Bayesian linear mixed-effects model to examine the effect of modal adjacency on the probability rating of nested epistemic expressions (see Mod 2 in Section 2.4). This model used word order (High-Low vs Low-High), the presence of the parenthetical element (present vs absent), and their interactions to predict the probability rating of nested expressions. The output of the model is summarized in Table 6. As what we found in the first experiment, there was no reliable effect of word order (β = 0.05, 95% CI = [-0.11, 0.22]) and there was no interaction between

the word order and the adjacency of the nested modals (β = 0.13, 95% CI = [-0.10, 0.36]). Moreover, in this experiment, the effect of the parenthetical element was not statistically meaningful for the nested modals with the High-Low word order, as the slope of the *With-Parenthetical* condition included the value of zero (β = -0.03, 95% CI = [-0.22, 0.15]).

Table 6. Experiment 2: The output of Mod 2, the effect of word order, modal adjacency, and their interaction on how people process nested epistemic expressions

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.86	0.12	0.62	1.10
Low-High	0.05	0.09	-0.11	0.22
With-Parenthetical	-0.03	0.09	-0.22	0.15
Low-High:With-Parenthetical	0.13	0.12	-0.10	0.36

[a]An estimate is statistically meaningful when zero is not included within the 95% credible interval. The baseline for comparison is the High-Low condition without a parenthetical element.

3.6. Discussion

We replicated the major findings of the first experiment using a modified paradigm that has a better ecological validity. Instead of reading the script of a dialogue containing epistemic expressions, participants in this experiment heard the dialogue unfolding the same way as they overheard a conversation between two interlocutors in daily situations. This created a more natural context for participants to process nested epistemic expressions in the dialogue.

Though small fluctuations were observed in the probability rating of a couple of modals, in general, the rating scores of various epistemic modals revealed a gradient scale. The difference in probability rating between the modals of higher epistemic strength (High condition) and lower epistemic strength (Low condition) was noticeable and such a difference held regardless of whether or not a parenthetical element was present. As we found in the first experiment, adding a parenthetical element boosted the probability rating of the statement embedded within a single modal of lower strength, but the opposite effect was found for statements embedded within modals of higher strength. A parenthetical element thus seems to function like a hedge, pulling the epistemic strength of a single modal from more extreme positions to the middle ground.

Importantly, the nested-modal data showed no effect of the word order regardless of the adjacency of the modals. In general, the mean rating score of the *High-Low* condition was slightly lower than that of the *Low-High* condition; however, such a small difference was not statistically meaningful. For the nested-modal data, the slope of the effect for the parenthetical element and its interaction with the word order had a credibility interval including zero. This suggested that changing the order of the modals in a nested expression would not change the perceived probability of the embedded statement, regardless of the adjacency of the two modals.

These findings again did not support the scope account, which predicted that the meaning of the nested expressions depends on the order of the two component modals. In Experiments 1 and 2, we used the strength of the epistemic expression as a proxy for its meaning, and found that for the nested epistemic expressions, changing the order of the component modals did not change how people interpreted the strength of the expression. It is likely that when interlocutors encounter an epistemic expression consisting of "non-harmonic" modals, they interpret such expression as an indicator

of medium degree of uncertainly in between the epistemic strength of the component modals. For example, given a statement "S", the expression of both "certainly may S" and "may certainly S" mean that the speaker conveys uncertainty over the probability of "S", the strength of which is estimated to be higher than "may" but lower than "certainly". This interpretation is consistent with the good-enough processing framework (Ferreira & Lowder, 2016).

In the previous two experiments, the semantic aspect we focused on was the degree of probability conveyed by the nested epistemic expressions. Though we didn't find any effect of word order on participants' understanding of the probability expressed, it is still possible that there are semantic aspects other than overall epistemic strength that are sensitive to the scope, and thus would be affected by the ordering of the nested modals. According to Moss (2015, p. 29), the semantics of epistemic expressions can be modeled as having mental committee members vote for the acceptance of a statement. For example, "certainly S" means every committee member accepts "S", while "may S" means that some committee members accept "S". Based on this framework, the semantics of a nested epistemic expression can be translated into a decision making scenario in which a committee votes for the acceptance of a statement that contains an epistemic modal. For example, the sentence "Tom certainly may have forgotten" corresponds to the scenario in which every committee member accepts that "Tom may have forgotten", while the sentence "Tom may certainly have forgotten" corresponds to a different scenario in which some committee members accept that "Tom certainly has forgotten". If we present participants with a nested epistemic expression and ask them to select among scenarios (similar to those listed above) the one most consistent with the expression, we should be able to examine the extent to which participants' interpretations of nested-epistemic expression match the prediction of the scope account. The experiment reported in the next section adopted this approach to explore the processing of nested epistemic expressions focusing on semantic aspects other than the overall epistemic strength.

4. Experiment 3

This experiment examined the scope account of processing nested epistemic expressions, testing whether or not the order of the nested modals affects how people interpret the meaning of the expression. Adopting the framework of mental committees (Moss, 2015), we asked participants to select the voting scenario that was best described by a statement. Importantly, some of the statements contained nested epistemic expressions, and some of the voting scenarios were supposed to reflect the semantics of those nested expressions based on the scope account. If participants constantly matched the statement to the target voting scenario predicted by the scope account, and also, the nested expressions in different word order were matched to different voting scenarios, that would serve as an evidence that the order of the nested epistemic expression affects how people interpret its meaning.

4.1. Participants

128 college students were recruited from the psychology research participation system at the University of California, Davis. They participated in the study in exchange for course credits. All participants were native speakers of English and naive concerning the purposes of the experiment. Following the pre-registered criteria of data inclusion

(ASPREDICTED #51982 available via `https://aspredicted.org/FNY_YLK`), participants who failed to achieve an accuracy rate above 90% for attention-checking filler items were excluded from analysis. As a result, data of 96 out of 128 participants were analyzed.

4.2. Stimuli

This study contained 48 experimental items. Each of the experimental items had a prompt describing a decision making context in which a group of 10 people voted to make a collective decision. The decision was described by a statement that contained epistemic modal(s), which fell into four different experimental conditions. As shown in Figure 7, the decision making context described in the prompt of this experimental item was 10 people voting for the best pie in a competition. The agreement that the committee reached was described by a statement which was manipulated in terms of the epistemic modal(s) it had. In the *High-Low* condition, the epistemic modal expressing higher probability (which is the word "certainly") preceded the modal expressing lower probability (which is the word "may"). In the *Low-High* condition, the epistemic modal expressing higher probability followed the modal expressing lower probability. Those two conditions were the nested-modal conditions since more than one epistemic modals occurred in the same clause. By comparison, the sentence in the single-modal conditions only contained one epistemic modal expressing either higher probability or lower probability.

NESTED-MODAL CONDITIONS	PROMPTS	OPTIONS
High-Low	*10 ladies on the church committee are voting on the best pie competition.* After deliberation, Mrs. Gherkin's blueberry pie **certainly may** win the purple ribbon.	**8 ladies** on the church committee agree they **may** vote for Mrs. Gherkin's blueberry pie.
		6 ladies on the church committee agree they are **certainly** going to vote for Mrs. Gherkin's blueberry pie.
Low-High	*10 ladies on the church committee are voting on the pie competition.* After deliberation, Mrs. Gherkin's blueberry pie **may certainly** win the purple ribbon.	**8 ladies** on the church committee agree they **may** vote for Mrs. Gherkin's blueberry pie.
		6 ladies on the church committee agree they are **certainly** going to vote for Mrs. Gherkin's blueberry pie.

SINGLE-MODAL CONDITIONS	PROMPTS	OPTIONS
High	*10 ladies on the church committee are voting on the best pie competition.* After deliberation, Mrs. Gherkin's blueberry pie is **certainly** going to win the purple ribbon.	**8 ladies** on the church committee are going to vote for Mrs. Gherkin's blueberry pie.
		6 ladies on the church committee are going to vote for Mrs. Gherkin's blueberry pie.
Low	*10 ladies on the church committee are voting on the best pie competition.* After deliberation, Mrs. Gherkin's blueberry pie **may** win the purple ribbon.	**8 ladies** on the church committee are going to vote for Mrs. Gherkin's blueberry pie.
		6 ladies on the church committee are going to vote for Mrs. Gherkin's blueberry pie.

For illustration purposes, in the prompt, the decision-making context was highlighted in italic, and the epistemic expressions were highlighted in bold. In the options, both the epistemic expressions and the number of the voters were written in bold font. None of the above information was highlighted in the actual experiment.

Figure 7. Experiment 3: Example of an experimental item.

After reading the prompt, participants were instructed to select from two options the one that was most consistent with the statement in the prompt. Those options described voting scenarios in which the number of the voters and/or the proposition they agreed upon differed slightly. Those options were constructed based on the mental

committee framework (Moss, 2015) to reflect the semantics of the epistemic expression in the statement. For an experimental item in the single-modal condition, one of the options was supposed to be semantically equivalent to the prompt in the *High* condition, while the other option was supposed to be semantically equivalent to the *Low* condition. In the example shown in Figure 7, the prompt of the *Low* condition, which said "Mrs. Gherkin's blueberry pie may win the purple ribbon", is most consistent with the voting scenario in which six ladies on the church committee voted for Mrs. Gherkin's blueberry pie. By comparison, the prompt of the *High* condition, "Mrs. Gherkin's blueberry pie is certainly going to win the purple ribbon", corresponded to the option saying "8 ladies on the church committee are going to vote for Mrs. Gherkin's blueberry pie". For the two options in the single-modal conditions, the number of voters reflected the epistemic strength of the modal in the prompt. For instance, the epistemic strength of the modal "certainly", as reflected from the rating of the previous two experiments, was about 80 out of 100, the ratio of which is equivalent to eight out of ten committee members. Similarly, the epistemic strength of the modal "may" on average was about 60 out of 100, which was represented by six out of ten committee members. A list of epistemic modals and its corresponding interpretation based on the framework of mental committees is shown in Figure 8.

MODALS	EPISTEMIC STRENGTH	MENTAL COMMITTEE INTERPRETATION
Certainly	80%	8 out of 10
Definitely	80%	8 out of 10
Must	80%	8 out of 10
Probably	70%	7 out of 10
May	60%	6 out of 10
Might	60%	6 out of 10
Possibly	60%	6 out of 10

Figure 8. Epistemic modals and its interpretation based on the framework of mental committees

The two options for the nested-modal conditions were more complicated, which contained the number of the voters and an embedded epistemic modal. For an experimental item in the nested-modal condition, one of the options was supposed to be semantically equivalent to the prompt in the *High-Low* condition, while the other option was supposed to be semantically equivalent to the prompt in the *Low-High* condition. According to the scope account, the meaning of the sentence in *High-Low* condition, "Mrs. Gherkin's blueberry pie certainly may win the purple ribbon", means "it is certainly the case that Mrs. Gherkin's blueberry pie may win the purple ribbon". This can be represented as "eight out of ten ladies agree that they may vote for Mrs. Gherkin's blueberry pie" based on the mental committee framework. Similarly, the prompt in the *Low-High* condition can be represented as "6 ladies on the church committee agree they are certainly going to vote for Mrs. Gherkin's blueberry pie". Participants were asked to pick the option that was most consistent with the prompt, and if participants constantly matched the prompt to the target voting scenarios predicted by the scope account, and also, the nested expressions in different word orders

were matched to different options, that would serve as an evidence that the order of the nested epistemic expression affected how people interpreted its meaning.

For the experimental item illustrated in Figure 7, the epistemic modals in question were "certainly" and "may", while other experimental items may have a different combination of epistemic modals. In this study, we selected six pairs of nested modals, which were "definitely and may", "definitely and might", "certainly and might", "certainly and may", "probably and might", and "probably and may". Each modal combination contained two epistemic modals, one expressing higher probability than the other, and each modal combination appeared in eight different experimental items.

The experimental items were assigned to 4 different lists, following a Latin Squared Design, and in total, there were 48 experimental items in each list. In addition to experimental items, each list also included 48 filler items. Each filler item also had a prompt describing a decision-making context and two options describing voting scenarios. Different from the experimental items, the filler item did not contain any epistemic modals, or it only contained an epistemic expression of negative probability, such as "impossible" or "unlikely". Among those filler items, 25 of them were attention-checking items in that only one option was logically reasonable. An example of the checking items is as below:

Prompt: A group of 10 friends are at Disneyland for the day. They're trying to decide which ride to go on first. After much deliberation, it's impossible that they'll go on Thunder Mountain first.
Option 1: 5 friends do want to ride Thunder Mountain first.
Option 2: 9 friends do not want to ride Thunder Mountain first.

It was explicitly stated in the prompt that it was impossible for the group to go on Thunder Mountain first, and based on that, we can reason that most of the people in the group voted against Thunder Mountain first, which was consistent with the second option rather than the first option. If a participant chose the first option, we would mark it as incorrect. If a participant's accuracy rate for all the attention-checking items was below 90%, the data of that participant would be excluded from analysis.

4.3. Procedure

Participants used personal computers to access the web page hosting this online study. We included an instruction with two practice trials before the experimental items. The practice trails mimicked the format of the experimental items in which there was a prompt followed by two options for participants to choose from. The prompt and the options were on the same page, and participants were instructed to select one of the two options that was most consistent with the information in the prompt. After the practice trials, participants hit a button to proceed. They were randomly assigned to one list of experimental items, and the first experimental item in the list was presented on the computer monitor screen. Participants read the prompt and selected the option the same way as they did in the practice trails. After that, they hit the proceed button to reveal the next experimental item. Once the new experimental item revealed itself, there was no way to return to the previous item. The study ended after the participant completed the task for all 96 the items in the list.

4.4. Data analysis

Participants' data in single-modal conditions and nested-modal conditions were analyzed separately. The options participants selected in the force-choice task were recoded in numerical values. For experimental items in single-modal conditions, we used number "1" to code the option that corresponded to the *High* condition and "0" to code the option that corresponded to the *Low* condition. For experimental items in nested-modal conditions, we used number "1" to code the option that corresponded to the *High-Low* condition and "0" to code the option that corresponded to the *Low-High* condition. We anticipated that for experimental items in single-modal conditions, participants were more likely to select "1" if they were shown a prompt in the *High* condition compared with seeing a prompt in the *Low* condition. Similarly, if the scope account predicted the way people processed nested epistemic expressions, participants would be more likely to choose "1" for prompts in the *High-Low* condition than for prompts in the *Low-High* condition.

Bayesian generalized linear models (GLM) were constructed using R package brms (Bürkner, 2017) with default priors [4] to explore the effects of modal arrangement on participants' choice preference. The probability of choosing "1" was modeled as a function of modal condition using logistic regression. Maximal random effects structures were constructed including subject and item intercepts and slopes, following the model structure below:

$$\begin{aligned} Choosing\ option\ 1 = 1 + condition \\ +(1 + condition|subject) \\ +(1 + condition|item) \\ family = bernoulli\ (link = \text{“}logit\text{”}) \end{aligned}$$

We built two models of the above structure for items in the single-modal conditions and double-modal conditions separately. The fixed effects were dummy-coded, with the *High* condition being the baseline for the single-modal conditions, and *High-Low* condition being the baseline for the nested-modal conditions. The data and script used for statistical analysis of this study is available in the GitHub[5].

4.5. Results

The frequency of each option selected by participants in different experimental conditions is shown in Figure 9. For experimental items that only contained a single epistemic modals, when participants were shown a prompt in the *High* condition, they selected the option corresponding to the semantics of the *High* modal in 1083 trials, while for the other 69 trials, they selected the option corresponding to the semantics of the *Low* modal. When participants were shown a prompt in the *Low* condition, they selected the option corresponding to the *High* modal in 587 of the trials, and selected the option corresponding to the *Low* modal in 565 of the trials.

The difference between *High* and *Low* conditions in the probability of choosing the option corresponding to the *High* modal is further illustrated in Figure 10(a) and

[4] Information on brms' default prior setting can be accessed following https://search.r-project.org/CRAN/refmans/brms/html/set_prior.html

[5] https://github.com/PON2020/Nested_Epistemic_Expressions_Further_Inquiry

	HIGH OPTION	LOW OPTION
High Prompt	1083	69
Low Prompt	587	565

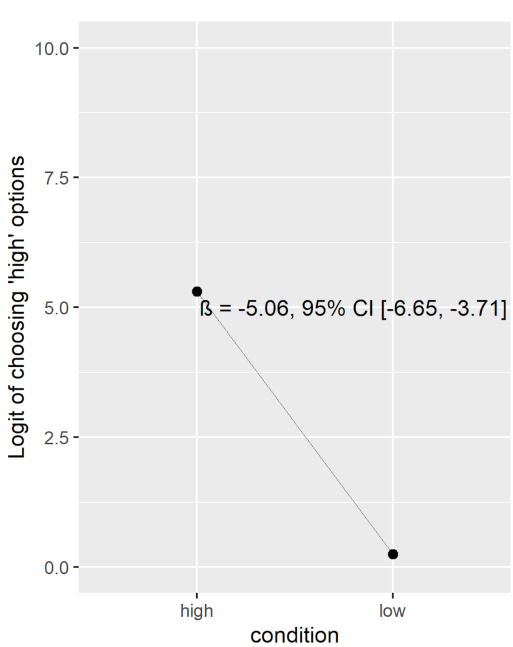
(a) Frequency of selecting each option in the single-modal conditions

	HIGH-LOW OPTION	LOW-HIGH OPTION
High-Low Prompt	766	386
Low-High Prompt	709	443

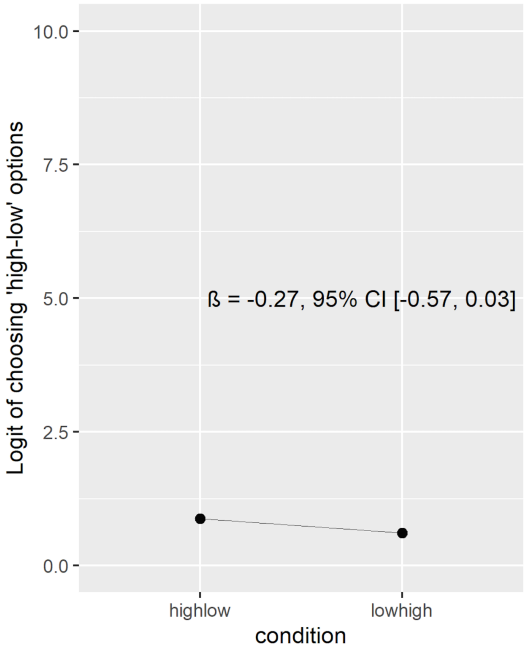
(b) Frequency of selecting each option in the nested-modal conditions

Figure 9. Experiment 3: Frequency of the options selected in different experimental conditions

Table 7. When reading the prompt that contained a single epistemic modal expressing lower probability, participants were less likely to select the option that was semantically equivalent to the *High* modal (β = -5.06, 95% CI = [-6.65, -3.71]) compared with the baseline condition in which the participants were shown a prompt that contained a single epistemic modal expressing higher probability. The difference between High and Low condition in the probability of choosing the *High* option is reflected in Figure 10(a) as a noticeable negative slope.



(a) Logit of selecting the option consistent with the prompt in the High condition



(b) Logit of selecting the option consistent with the prompt in the High-Low condition

Figure 10. Experiment 3: Fitting the logistic regression models for the single and nested conditions

For the experimental items that contained nested epistemic expressions, when participants were shown a prompt that contained two nested modals in the *High-Low*

order, they selected the option corresponding to the semantics of the *High-Low* modal combination in 766 trials, while for the other 386 trials, they selected the option corresponding to the semantics of the *Low-High* combination. When participants were shown a prompt that contained two nested modals in the *Low-High* order, they selected the option corresponding to the *High-Low* combination in 709 trials, and selected the option corresponding to the *Low-High* combination in 443 trials. The output of the logistic regression showed that the probability of selecting the option semantically equivalent to the *High-Low* expression remained nearly the same regardless of whether the participants were given a prompt containing *High-Low* combination or *Low-High* combination (see Figure 10(b)). When participants were shown a prompt containing a nested epistemic expression in *Low-High* word order, they were slightly less likely to select the option that was semantically equivalent to the *High-Low* modal combination, compared to the baseline condition in which a *High-Low* prompt was given. However, such a small difference between the *High-Low* condition and *Low-High* condition was not statistically meaningful, given that zero was included in the credibility interval of the slope (β = -0.27, 95% CI = [-0.57, 0.03, see Table 8).

Table 7. Experiment 3: Estimate of intercept and slopes for single-modal data with the High condition as the baseline

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	5.31	0.69	4.14	6.78
Low	-5.06	0.75	-6.65	-3.71

[a]An estimate is statistically meaningful when zero is not included within the 95% credible interval.

Table 8. Experiment 3: Estimate of intercept and slopes for nested-modal data with the High-Low condition as the baseline

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.88	0.15	0.58	1.19
Low-High	-0.27	0.15	-0.57	0.03

[a]An estimate is statistically meaningful when zero is not included within the 95% credible interval.

4.6. Discussion

In this experiment, we investigated the processing of nested epistemic expressions by creating stories that would elicit different interpretations of the expression and asking participants to select the interpretation that best represents their own understanding of the expression they read. Among the options, there was one interpretation that was supposed to be the correct interpretation based on the scope account. If participants constantly selected the theoretically preferred interpretation, and also if the difference in the word order led to a change in the preferred option, we would be confident that the underlying mechanism of processing nested epistemic expression is in line with the scope account.

Based on the mental committee framework, for statements containing only one epistemic modal, the strength of that modal is reflected from the proportion of the committee members who vote for a certain proposal. For example, "Mrs. Gherkin's blueberry pie is certainly going to win the purple ribbon" means eight out of ten ladies on the church committee are going to vote for Mrs. Gherkin's blueberry pie, while " Mrs. Gherkin's blueberry pie may win the purple ribbon" means six out of ten ladies on the church committee are going to vote for Mrs. Gherkin's blueberry

pie. The probability expressed by the modal "certainly" is higher than the probability expressed by the modal "may", and such a difference is represented by more votes in the scenario that corresponds to the semantics of "certainly". The result of this experiment showed that participants were more likely to match the scenario "8 ladies on the church committee are going to vote for Mrs. Gherkin's blueberry pie" to the statement "Mrs. Gherkin's blueberry pie is certainly going to win the purple ribbon" rather than to the statement "Mrs. Gherkin's blueberry pie may win the purple ribbon". This suggested that the semantic difference between the single modal "certainly" and "may" can be represented using the mental committee framework, and such a representation was in line with how participants processed the meaning of the modal "certainly" and "may". The next question is whether or not participants' processing of nested epistemic expressions is also consistent with the prediction of the mental committee framework.

To create the mental committee representations of expressions containing two epistemic modals, we used the number of votes to represent the outer modal, while the inner modal was embedded in the proposal. Based on this account, the statement "Mrs. Gherkin's blueberry pie certainly may win the purple ribbon" is semantically equivalent to the option "8 ladies on the church committee agree they may vote for Mrs. Gherkin's blueberry pie", while "Mrs. Gherkin's blueberry pie may certainly win the purple ribbon" means "6 ladies on the church committee agree they are certainly going to vote for Mrs. Gherkin's blueberry pie". What we found in this experiment was that the chance of selecting the option "8 ladies on the church committee agree they may vote for Mrs. Gherkin's blueberry pie" remained the same regardless of whether the prompt the participants saw was "Mrs. Gherkin's blueberry pie certainly may win the purple ribbon" or "Mrs. Gherkin's blueberry pie may certainly win the purple ribbon". This suggested that the order of the modals in a nested epistemic expression had no effect on how participant interpreted the meaning of such expression, which was not consistent with the prediction under the scope account.

Similar to the previous experiments, the findings of this experiment is more in line with the prediction of the good-enough processing framework (Christianson et al., 2001, 2006; Ferreira et al., 2002, 2001; Ferreira & Lowder, 2016). According to this account, the parser performs superficial analysis of linguistic input based on heuristics, leading to inaccurate interpretations. In terms of the processing of nested epistemic expressions, the scope relation between the nested modals is not being processed in casual conversation, rather, interlocutors treat the occurrence of more than one epistemic modals with contrasting epistemic strength as an indication of uncertainty over the topic being discussed. Thus, the order in which the modals occur in the utterance does not affect how participants interpret the meaning of the nested epistemic expression.

5. General discussion

Though the semantics of epistemic modals has been extensively elaborated in the theoretical literature, few attempts have been made to account for the cognitive mechanisms in which people process nested epistemic expressions. In three experiments, we examined the processing of nested epistemic modality, focusing on how interlocutors interpret modally non-harmonic expressions like "He certainly may have forgotten". Based on the scope account, interlocutors interpret the meaning of the second modal within the scope of the first modal (Moss, 2015; Potsdam, 1998), and thus, the mean-

ing of the nested expression depends on the order in which the two modals occur. On the other hand, the good-enough processing account (Christianson et al., 2001, 2006; Ferreira et al., 2002, 2001; Ferreira & Lowder, 2016) suggests the possibility that nested epistemic expressions are processed in a shallow manner, and thus, the order of the two modals does not affect how interlocutors interpret the meaning of the nested expression. Qiu and Ferreira (2022) tested the above accounts by asking participants to rate the probability of the statements embedded in nested epistemic expressions while the order of the modals in the nested expressions was manipulated. What they found was that the order of the component modals had no effect on the perceived strength of the nested expression. This pattern is more consistent with the prediction of the good-enough processing framework rather than the scope account. Qiu and Ferreira (2022) sets the stage for systematic investigation of processing nested epistemic expressions.

The experiments reported in this paper examined the findings of Qiu and Ferreira (2022) and further explored questions that have not been fully answered. To start with, in this study, the effect of word order was evaluated with regard to a second factor, the adjacency of the two nested modals. It is possible that the scope of the nested epistemic expressions is more salient when the two component modals are separated rather than adjacent, and if that is the case, we are expected to see an interaction between the order and the adjacency of the nested modals. To evaluate this possibility, the first experiment adopted a paradigm similar to that of Qiu and Ferreira (2022) while including both word order and the adjacency of the modals in a within-participant design. The second experiment replicated the first experiment with all the stimuli presented auditorily rather than visually, a set-up that closely resembled casual conversations. We believe that Experiment 1 and 2 together would provide a solid account for how people interpret the overall strength of nested epistemic expressions. What we found in both Experiments 1 and 2 was that the order of the component modals had no effect on how people interpret the strength of the nested expressions, regardless of whether the two modals were adjacent or separated by a parenthetical element. Experiment 3 focuses on the semantic aspect other than the overall epistemic strength. By providing participants with different interpretations of a nested epistemic expression and asking them to select the one closer to their own interpretation, Experiment 3 examined the mental committee framework as an example of the scope account, but still, the order of the modals did not affect how nested epistemic expressions were interpreted. Thus, the findings of all three experiments were not consistent with the claim that interlocutors interpret the meaning of a nested epistemic expression by analyzing one modal within the scope of the other modal.

Scope is undoubtedly a crucial construct indispensable to successful language comprehension. It describes the relative priority of different logical operations, from which the processor derives the syntactic and semantic relationship between the constituents in an utterance (Hintikka, 1997; Ladusaw, 1979). The past century has witnessed lively discussions of the cognitive mechanisms underlying the processing of quantifier scopes (see Kurtzman and MacDonald, 1993 for a review), and some general principles have been proposed. First of all, it is very common that the ordering of the scopes follows the linear order of the phrases in the surface structure of the sentence (Fodor, 1982; Johnson-Laird, 1969). As previously illustrated, the scope of the negation "not" is wider in the sentence "Not every student is happy" compared with that in the sentence "Every student is not happy", and such a difference is revealed from the ordering of the negation and the universal quantifier. On the other hand, many believe that it is not the surface word order that determines scope relations; rather, the scopes

of the constituents are processed according to their positions in the syntactic hierarchy (Ladusaw, 1979; Reinhart, 1983). For example, the scope relation between the existential quantifier "a" and the universal quantifier "every" remains the same in the sentence "A kid climbed every tree." and "Every tree was climbed by a kid." though the linear order of the two quantifiers differs in the surface structure of the sentence. The reason is that the active sentence and its passive counterpart share the same syntactic "deep structure" (Chomsky, 1971), which determines how the scopes of the operators are processed. Moreover, some processing heuristics have been proposed which are less "syntactically analytical" in nature, for example, the phrases corresponding to the sentence topic tend to have a wider scope (Kempson & Cormack, 1981); semantic agent tends to have a wider scope than a theme (Grimshaw, 1990). With regard to all these principles mentioned above, Kurtzman and MacDonald (1993) observed that 1) More than one principle is at play during the online processing of quantifier scopes, and the preferred interpretation is probabilistic in nature. 2) For active sentences in English, the linear order of the constituents predicts how their scopes are processed: the leftward quantifiers are interpreted as having wider scopes (1993:257).

The epistemic modals we discussed in this paper express quantification over possible worlds (Alechina et al., 1995; Kratzer, 2012), and the processing of an epistemic expression involves identifying the scope of the modal relative to the embedded statement. Computation then takes place assigning to the embedded statement the speaker's estimate of its truth value as indicated by the strength of the epistemic modal. For example, the expression "Tom may have forgotten the meeting" is semantically equivalent to "It may be the case that Tom has forgotten the meeting." In both cases, comprehenders assign a low probability to Tom's forgetting of the meeting. Similarly, the semantics of an expression containing multiple epistemic modals has been argued to involves the computation of nested scopes (Lyons, 1977; Moss, 2015; Potsdam, 1998), and that is why the term "nested epistemic expressions" has been given to sentences like "Tom certainly may have forgotten". As suggested by the scope account, the above expression is semantically equivalent to "it is certainly the case that it may be the case that Tom has forgotten." This expression assigns a high probability to the statement "Tom may has forgotten", which in itself contains an epistemic modal and another statement within its scope.

Semanticians in general agree that a speaker's selection of an epistemic modal in the utterance is based on the evaluation of the factuality of the statement by considering all the evidence at hand. That is why the set of evidence is termed "modal base" in Kratzer (2012). If our previous knowledge of Tom consists of only some incidents of him being forgetful, that knowledge serves as the base for the utterance "Tom may have forgotten". On the other hand, if our previous knowledge of Tom is full of incidents of him being forgetful, the use of the modal "may" is not warranted. "Tom certainly has forgotten" is thus a more reasonable utterance in this case. An important argument of the scope account is that the base for the outer modal in a nested expression is the evidence concerning what is projected from the inner modal (Moss, 2015). For example, the base for the modal "certainly" in the utterance "Tom certainly may have forgotten" is the evidence explaining why it is highly likely that Tom may have forgotten. It is easy to argue why it is "certain" that Tom "may" have forgotten: Every human-being "may" be forgetful. It is certainly the case that Tom may be forgetful, because Tom is a human-being. Following the same line of reasoning, the scope account interprets "Tom may certainly have forgotten" as saying "it is possible that Tom certainly has forgotten". In this case, the base of the modal "may" should explain why it is possible that Tom is very likely to forget. All in all,

the scope account for the processing of nested epistemic expressions is analogous to the mechanisms that have been proposed to account for the processing of quantifier scopes: the linear order of the modals predicts how their scopes are processed, and the truth conditions differ for the nested expressions with different orders of embedding.

The good-enough account of language processing does not deny that comprehenders have the competence to assign different truth conditions to "Tom certainly may have forgotten" and "Tom may certainly have forgotten", specifically if the scope difference of the two modals is highlighted with a more "expanded" sentence structure, such as "It is certainly the case that Tom may have forgotten" vs "It is possibly the case that Tom certainly has forgotten". What we argue in this paper is that in casual communications where nested epistemic expressions most likely to appear, interlocutors tend not to understand the scope of the two modals as "nested", rather, they adopt a heuristic algorithm which features fast but shallow processing. Cases of fast and shallow processing abound in the psycholinguistic literature. At the lexical level, the processor becomes less stringent in detecting semantic anomalies when it has a strong prior belief of the context, leading to semantic illusions such as "bury the survivors" (Barton & Sanford, 1993) and the famous "Moses illusion" (Erickson & Mattson, 1981). At the discourse level, readers may normalize the text they read by neglecting the contradictory information it contains and interpreting the text as a coherent piece (Otero & Kintsch, 1992).

At the syntactic level, we see evidence that the parser may fail to construct the syntactic representation that is faithful to the input, especially when the prior knowledge gives less support to the input. For example, implausible sentences like "The mother gave the candle the daughter." is frequently interpreted as "The mother gave the candle to the daughter." (Gibson et al., 2013), and a recent study found evidence that the syntactic structure corresponding to the non-literal interpretation of the ditransitive has been constructed (Cai, Zhao, & Pickering, 2022). Ferreira and Lowder (2016, p. 222) noted that the representation derived from the syntactic algorithm is influenced by the representation derived from heuristics, leading to a complex state of knowledge with literal and non-literal representations. For example, even though the parser has recovered from the garden path effect, analyzing the sentence "While Mary bathed the baby played in the crib." correctly as saying the baby played in the crib while Mary bathed herself, it nevertheless preserves the representation of the incorrect parse, which is "Mary bathed the baby".

All these findings suggest the possibility that the nested epistemic expressions in which two modals express contradictory epistemic strengths could also undergo certain normalization processes in daily communication, leading to comprehenders not forming a veridical internal representation of the linguistic input they received (Traxler, 2014). A possible mechanism is that when processing non-harmonic expressions, the strength of the first modal lingers in memory, and is mixed with the epistemic strength of the second modal when it is encountered. The processor then normalizes the mixed epistemic strength as an indicator of general uncertainty, and assigns an intermediate degree of probability to the expression without further pinning down the scope of each modal. Thus, it is not likely for comprehenders to infer two different sets of evidence (model bases) from the use of two modals in a nested expression. Instead, comprehenders treat the two modals together as indicating an overall lack of certainty, and infer that there must be a reason for the speaker to be unsure. This mechanism underlies the processing of nested epistemic expressions regardless of whether the component modals are adjacent or separated by other words in between.

Finally, we would like to briefly elaborate on an interesting observation we made in

our first two experiments, in which the presence of a parenthetical element was an independent variable manipulating the distance of the modals. Though the focus of this project is on the nested modals, we nevertheless found an interaction effect between the presence of a parenthetical element and epistemic strength in expressions that contain only one epistemic modal. To be specific, we found that adding a parenthetical element boosted the probability rating of the statement embedded within the scope of a single *Low* modal; however, it decreased the probability rating of the statement embedded within the scope of a *High* modal. The parenthetical elements used in the first two experiments were evidential expressions indicating the source of information. Evidentiality is a notional category that has been grammaticalized in about one fourth of the world's languages (Aikhenvald, 2004, p. 17), often as inflectional morphemes (Chafe & Nichols, 1986). For example, in Tariana (an Arawakan language), the verb "play" in the expression "Jose has played football" is attached with different inflectional morphemes, depending on whether the speaker has seen Jose play football, or heard that Jose play football (Song, 2018). English does not have grammaticalized evidentials, and the notion of evidentiality is often expressed by using parenthetical elements such as "based on what I saw" and "based on what I heard" etc. Reinhart (1983, p. 175) suggested that parenthetical evidentials in English can often be regarded as an epistemic expression. The example she gave was the utterance "Jules will be late, he said", in which, the evidential "he said" modifies the strength of "Jules will be late" similar to a typical epistemic modal like "probably" (as in "Jules will probably be late"). If Reinhart (1983) is correct in saying the function of parenthetical evidentials in English is similar to epistemic modals, the expression containing an epistemic modal and an evidential can be regarded as a nested epistemic expression. Based on the mechanism we proposed, the processor derives an intermediate overall epistemic strength when processing nested epistemic expressions, and that explains why parenthetical evidentials pull the strength of a single modal from more extreme positions to the middle ground.

6. Conclusion

The findings of our experiments provide answers to the research questions listed in the Introduction section. In terms of the compositionality of the nested epistemic expression, both of the component modals contribute to the meaning of the nested expression. Specifically, when the component modals differ in their epistemic strength, the overall strength of the nested expression lies in between the epistemic strength of the component modals. As to the scope of the modals, although some formal theories stipulate that the inner modal should be interpreted within the scope of the outer modal, the data from our experiments do not support this argument. We argue that for nested epistemic expressions in casual conversations, interlocutors often do not process the scope of the modals, rather, they interpret the presence of two epistemic modals as indicating lack of certainty, and thus, the order of the nested modals is free to vary without changing interlocutors' understanding of the expression. Moreover, the adjacency of the two modals do not influence how the nested expressions are processed. The findings of this research challenge the scope account of nested epistemic expressions in language processing (Moss, 2015; Potsdam, 1998), and suggest a holistic processing mechanism in line with the "good enough" processing framework (Christianson et al., 2001, 2006; Ferreira et al., 2002, 2001; Ferreira & Lowder, 2016).

Acknowledgments

We thank Dr. Kenji Sagae for his insightful comments on this project. We thank all the undergraduate research assistants who helped in conducting our experiments.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Additional information

Funding

This work was supported by National Science Foundation Grant BCS-1650888 and R01HD100516-01A1 (to Fernanda Ferreira) and a UC Davis Hellman Fellowship (to Emily Morgan).

References

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford University Press.

Alechina, N. A., et al. (1995). *Modal quantifiers*. Institute for Logic, Language and Computation.

Barton, S. B., & Sanford, A. J. (1993). A case study of anomaly detection: Shallow semantic processing and cohesion establishment. *Memory & cognition*, *21*(4), 477–487.

Boersma, P. (2011). Praat: doing phonetics by computer [computer program]. *http://www. praat. org/*.

Bürkner, P.-C. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of statistical software*, *80*(1), 1–28.

Bybee, J. L. (1985). *Morphology: a study of the relation between meaning and form* (No. v. 9). Amsterdam ; Philadelphia: J. Benjamins.

Cai, Z. G., Zhao, N., & Pickering, M. J. (2022). How do people interpret implausible sentences? *Cognition*, *225*, 105101.

Chafe, W. L., & Nichols, J. (Eds.). (1986). *Evidentiality: the linguistic coding of epistemology* (No. v. 20). Norwood, N.J: Ablex Pub. Corp.

Chomsky, N. (1971). Deep structure, surface structure, and semantic interpretation. *Semantics*, 183–216.

Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive psychology*, *42*(4), 368–407.

Christianson, K., Williams, C. C., Zacks, R. T., & Ferreira, F. (2006). Younger and older adults'" good-enough" interpretations of garden-path sentences. *Discourse processes*, *42*(2), 205–238.

Coates, J. (1983). *The semantics of the modal auxiliaries*. Routledge.

Degen, J., Trotzke, A., Scontras, G., Wittenberg, E., & Goodman, N. D. (2019). Definitely, maybe: A new experimental paradigm for investigating the pragmatics of evidential devices across languages. *Journal of Pragmatics*, *140*, 33–48.

Di Paolo, M. (1989). Double modals as single lexical items. *American Speech*, *64*(3), 195–224.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540–551.

Ferreira, F. (2003). The misinterpretation of noncanonical sentences. *Cognitive psychology*, *47*(2), 164–203.

Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current directions in psychological science*, *11*(1), 11–15.

Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of garden-path sentences: Implications for models of sentence processing and reanalysis. *Journal of psycholinguistic research*, *30*(1), 3–20.

Ferreira, F., & Lowder, M. W. (2016). Prediction, information structure, and good-enough language processing. In *Psychology of learning and motivation* (Vol. 65, pp. 217–247). Elsevier.

Fodor, J. D. (1982). The mental representation of quantifiers. *Processes, beliefs, and questions: Essays on formal semantics of natural language and natural language processing*, 129–164.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325.

Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive psychology*, *14*(2), 178–210.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*(20), 8051–8056.

Grimshaw, J. (1990). *Argument structure.* the MIT Press.

Halliday, M. A. (1970). Functional diversity in language as seen from a consideration of modality and mood in english. *Foundations of language*, 322–361.

Halliday, M. A., & Matthiessen, C. (2004). *An introduction to functional grammar* (3rd ed ed.). London : New York: Arnold ; Distributed in the United States of America by Oxford University Press.

Hintikka, J. (1997). No scope for scope? *Linguistics and Philosophy*, 515–544.

Holmes, J. (1982). Expressing doubt and certainty in english. *RELC journal*, *13*(2), 9–28.

Huang, Y., & Ferreira, F. (2021). What causes lingering misinterpretations of garden-path sentences: Incorrect syntactic representations or fallible memory processes? *Journal of Memory and Language*, *121*, 104288. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0749596X21000711`

Huddleston, R., & Pullum, G. K. (2002). *The cambridge grammar of the english language.* Cambridge University Press.

Hyland, K., & Milton, J. (1997). Qualification and certainty in l1 and l2 students' writing. *Journal of second language writing*, *6*(2), 183–205.

Johnson-Laird, P. N. (1969). On understanding logically complex sentences. *Quarterly Journal of Experimental Psychology*, *21*(1), 1–13.

Kempson, R. M., & Cormack, A. (1981). Ambiguity and quantification. *Linguistics and Philosophy*, *4*, 259–309.

Kratzer, A. (2012). The Notional Category of Modality. In *Modals and Conditionals.* Oxford: Oxford University Press.

Kurtzman, H. S., & MacDonald, M. C. (1993). Resolution of quantifier scope ambiguities. *Cognition*, *48*(3), 243–279.

Ladusaw, W. A. (1979). *Polarity sensitivity as inherent scope relations.* The University of Texas at Austin.

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge University Press.

Moss, S. (2015). On the semantics and pragmatics of epistemic vocabulary. *Semantics and Pragmatics*, *8*, 5–1.

Nagle, S. J. (1994). The english double modal conspiracy. *Diachronica*, *11*(2), 199–212.

Nagle, S. J. (2012). Double modals in the southern united states: Syntactic structure or syntactic structures? In R. Facchinetti, F. Palmer, & M. Krug (Eds.), *Modality in contemporary english* (pp. 349–372). De Gruyter Mouton. Retrieved from `https://doi.org/10.1515/9783110895339.349`

Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*, *3*(4), 229–236.

Perkins, M. R. (1983). *Modal expressions in english / michael r. perkins* [Book]. F. Pinter London. Retrieved from `http://www.loc.gov/catdir/toc/fy0605/88202602.html`

Potsdam, E. (1998). A syntax for adverbs. In *The proceedings of the twenty-seventh western conference on linguistics* (pp. 397–411).

Qiu, Z., & Ferreira, F. (2022). "he may certainly have forgotten": Processing of nested epistemic expressions. *Discourse Processes*, 1–28.

Reinhart, T. (1983). Point of view in language: The use of parentheticals. *Essays on deixis*, *188*, 169–194.

Renooij, S., & Witteman, C. (1999). Talking probabilities: communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning*, *22*(3), 169–194.

Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., & Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, *69*(2), 104–120.

Song, J. J. (2018). *Linguistic typology*. Oxford University Press.

Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, bayesian estimation, and good-enough parsing. *Trends in cognitive sciences*, *18*(11), 605–611.

Verkuilen, J., & Smithson, M. (2012). Mixed and mixture regression models for continuous bounded responses using the beta distribution. *Journal of Educational and Behavioral Statistics*, *37*(1), 82–113.

Willems, S. J., Albers, C. J., & Smeets, I. (2019). Variability in the interpretation of dutch probability phrases-a risk for miscommunication. *arXiv preprint arXiv:1901.09686*.

Appendix

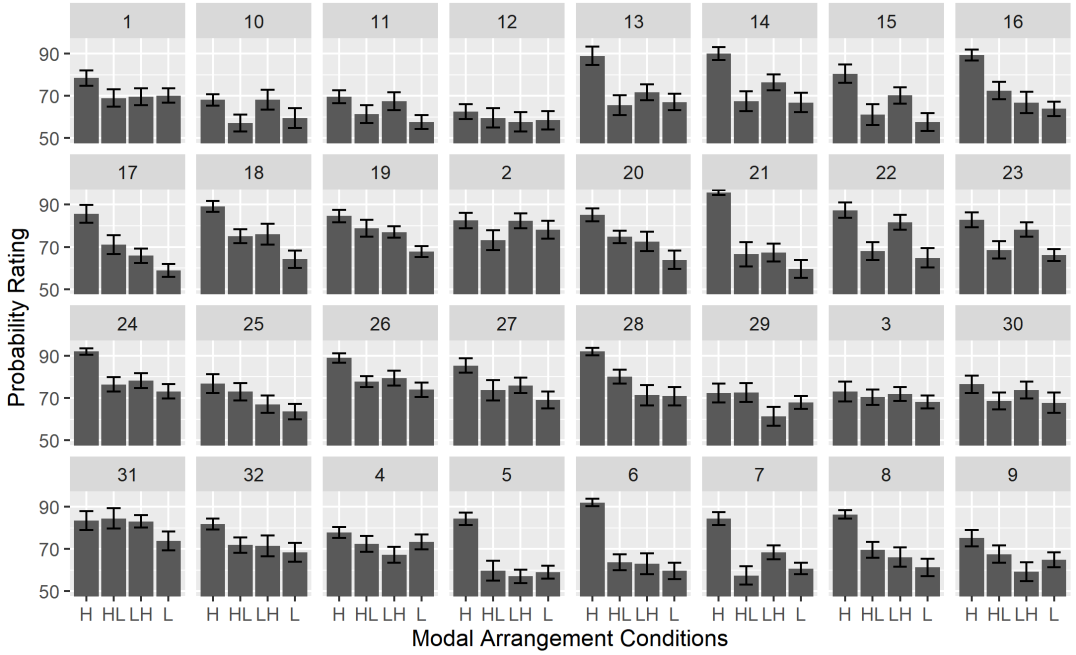


Figure 11. Experiment 1: Mean probability ratings of different modal arrangement conditions for each experimental item (with standard errors). The ratings were averaged for each experimental item across participants regardless of whether or not a parenthetical element was presented.

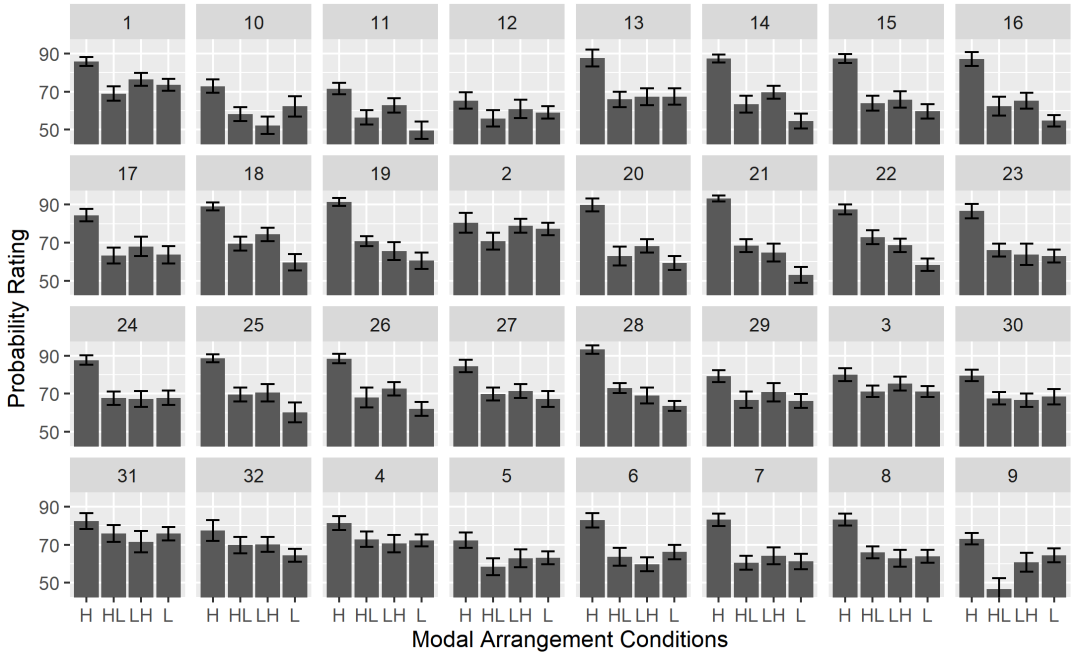


Figure 12. Experiment 2: Mean probability ratings of different modal arrangement conditions for each experimental item (with standard errors). The ratings were averaged for each experimental item across participants regardless of whether or not a parenthetical element was presented.