Discourse Processes

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hdsp20

"He May Certainly Have Forgotten": Processing of Nested Epistemic Expressions

Zhuang Qiu & Fernanda Ferreira

To cite this article: Zhuang Qiu & Fernanda Ferreira (2022) "He May Certainly Have Forgotten": Processing of Nested Epistemic Expressions, Discourse Processes, 59:8, 591-618, DOI: 10.1080/0163853X.2022.2077064

To link to this article: https://doi.org/10.1080/0163853X.2022.2077064

Published online: 06 Jun 2022.

DISCOURSE PROCESSES
2022, VOL. 59, NO. 8, 591–618
https://doi.org/10.1080/0163853X.2022.2077064

Routledge
Taylor & Francis Group

"He May Certainly Have Forgotten": Processing of Nested Epistemic Expressions

Zhuang Qiu [a] and Fernanda Ferreira [b]

[a]Department of Linguistics, University of California; [b]Department of Psychology, University of California

ABSTRACT

This article presents a series of three experiments investigating the processing of nested epistemic expressions, utterances containing two epistemic modals in one clause, such as "he *certainly may* have forgotten." While some linguists claim that in a nested epistemic expression one modal is semantically embedded within the scope of the other modal based on the word order, it is possible that in daily conversation the scope of nested modals may not be thoroughly processed, leading to a "good-enough" interpretation that is not sensitive to the word order of the two modals. This study used probability judgment tests to investigate people's interpretation of nested epistemic expressions, and the effect of word order was not observed. This result fails to support the scope account of the nested epistemic expressions and suggests a holistic processing mechanism in line with the good-enough processing framework.

Introduction

Suppose you plan to go hiking tomorrow and hear your friend say "it *might* be raining tomorrow." Will you bother canceling the trip? What if instead of hearing "it *might* be raining tomorrow," you hear a more confident assertion from that friend, saying "it is *certainly* going to rain tomorrow." Does it make a difference now? It has been claimed that humans often think and behave according to what things might be like, and the world view of uncertainty and probability forms "an essential part of the fabric of our everyday lives" (Perkins, 1983, p. 6). In the field of linguistics and logic, words such as *might* and *certainly* are referred to as *epistemic modals*, which indicate speakers' commitment to the truth value of what is said (Coates, 1983; Kratzer, 2012), and serve as an important means to modify the strength of an argument (Hyland & Milton, 1997).

A core dimension in the meaning of epistemic modals is "the strength of commitment to the factuality or actualisation of the situation," which is termed *strength of modality* by Huddleston and Pullum (2002, p. 175). For example, the speaker saying "it is *certainly* going to rain tomorrow" shows a stronger commitment to the coming rain than the speaker saying "it *might* rain tomorrow." In this way, the word *certainly* has a higher *epistemic strength* than the word *might*. Speakers' strength of commitment is influenced by the strength of evidence that supports the statement (Degen et al., 2019), and the spectrum of the epistemic strength has been categorized by many linguists using a three-point scale from high strength to medium strength to low strength (Halliday & Matthiessen, 2004; Holmes, 1982; Huddleston & Pullum, 2002). Halliday and Matthiessen (2004), for example, specified the lexical items under each of three categories: *certain (certainly)* and *must* express high degree of probability, *probable (probably)* and *will (would)* express medium probability, and *possible (possibly)* and *may (might)* express low probability. Other researchers, by contrast, preferred to represent epistemic

CONTACT Zhuang Qiu zkqiu@ucdavis.edu Department of Linguistics, University of California, 469 Kerr Hall, One Shields Avenue, Davis, CA, 95616

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

strength using a continuous scale rather than an ordinal representation (Degen et al., 2019; Renooij & Witteman, 1999; Willems et al., 2019). We will see in this study that both ordinal and continuous representations of epistemic strength served as important components in the research paradigm we adopted to study semantic processing of epistemic modality.

An interesting phenomenon about epistemic modals that has not been explicitly studied is the use of more than one epistemic modal in a single clause. For instance, the sentence "He may certainly have forgotten" has two modals, which are the words *may* and *certainly*. Following Moss's (2015) analysis of "nested epistemic vocabulary," we name the usage of double epistemic modals a *nested epistemic expression*. A closer look at the two modals in a nested epistemic expression reveals a semantic conflict in terms of the degree of probability expressed. While the word *may* indicates low probability, the word *certainly* expresses a high probability. That is why this kind of nested expression is called "modally nonharmonic" combination as opposed to "modally harmonic expression," like "he may possibly have forgotten," in which both the words *may* and *possibly* express low probability (Lyons, 1977).

Though the use of nested epistemic expressions is not common in formal registers, colloquial and informal communication has witnessed more frequent occurrence of nested expressions. To obtain a preliminary understanding of how frequently people use nested epistemic expressions in daily communication, we created a database of 413,986 tweets, each containing at least one epistemic modal, and searched for nested epistemic expressions in the database. In total, we found about 4,200 tweets containing nested epistemic expressions, with half of them being harmonic expressions like *may possibly* and the other half being nonharmonic expressions like *definitely might*. The script and report of this corpus study can be accessed from the project GitHub repository.[1] Based on this preliminary corpus search, it is estimated that there is roughly one case of nested epistemic expression out of a hundred cases of epistemic expressions. It is important to note that the two epistemic modals in a nested expression do not have to be adjacent to each other (e.g. "*Certainly* the candidate *might* win the election"), while the corpus analysis mentioned above only looked for the cases where the modals were adjacent. Thus, the frequency of nested epistemic expressions is likely higher than the above estimate.

Research on the processing of nested epistemic expressions sheds lights on our understanding of how the processor analyzes the scope of linguistic inputs 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, *a country greenhouse* means something different from *a green country house*. In the first case, the word *country* has within its scope a two-word compound *green-house*, meaning the greenhouse is in the country. In contrast, for the latter case, *country house* is within the scope of *green*, meaning the country house is green. This example shows that in order to derive a meaningful semantic interpretation of the linguistic input, the comprehender of English needs to sort out the scope of different logical operations indicated by the word order. Moreover, the notion of scope can be defined from both syntactic and semantic perspectives (Ladusaw, 1979). On one hand, we can define scope as a relation between constituents in the syntactic structure, and in that case, a constituent B is in the scope of constituent A, if and only if A c-commands B (Ladusaw, 1979, p. 37).[2] On the other hand, we can define scope as a relation between the meanings of two constituents when interpreting that piece of utterance. In that case, the scope of an expression A is "the constituent whose meaning is the argument of the meaning of A" (Ladusaw, 1979, p. 50). The notion of syntactic scope and semantic scope are not irrelevant. In fact, as De Swart (1998) noted, in English, the semantic scope is constrained by syntactic scope in that "the semantic scope of an operator involves at least its c-command domain" (p. 177).

With regard to nested epistemic expressions, it has been argued that one modal must be interpreted within the scope of the other modal (Lyons, 1977; Moss, 2015). Based on the syntactic structure of modal adverbs proposed by Potsdam (1998), the word *certainly* in the sentence "he certainly may have forgotten" is higher in the syntactic tree structure, c-commanding the modal auxiliary *may*. Thus, for

sentence "he certainly may have forgotten," the modal auxiliary *may* is within the scope of *certainly*. On the other hand, the word *certainly* in the sentence "he may certainly have forgotten" is c-commanded by the modal auxiliary *may*. Thus, in the sentence "he may certainly have forgotten," *certainly* is within the scope of *may*. As we can see, the scope of the modals in a nested expression is not only indicated by the word order, but also their relative position in the hierarchical structure of the syntax. Based on this account, the modally nonharmonic expression "He certainly may have forgotten" means something like *it is certainly the case that he may have forgotten*. Notice that for this interpretation, 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*.

Linguistic theories of scope draw a distinction between the meaning of "He certainly may have forgotten" and the meaning of "He may certainly have forgotten" (Moss, 2015; Potsdam, 1998), though the two utterances are identical except for the order of the two epistemic modals. The question is, when these nested epistemic expressions are encountered in daily communication, do interlocutors pin down the scope difference implied by the word order of the modals and assign different meanings to cases like the above? Some early studies on reading comprehension discovered that comprehenders frequently normalize the text they read, leading to a mental representation that is not entirely faithful to the content of the input (Barton & Sanford, 1993; Erickson & Mattson, 1981; Otero & Kintsch, 1992). Otero and Kintsch (1992), for example, asked participants to read paragraphs that contained contradictory statements. They found that many participants neglected the contradictory information and interpreted the text they read as a coherent piece. This finding suggested the possibility that the nested epistemic expressions in which two modals express contradictory epistemic strengths could also undergo certain normalization process in daily communication. If that is the case, during the processing of nested epistemic expressions, comprehenders might not form a veridical internal representation of the linguistic input they received (Traxler, 2014).

The good-enough theory of language processing (Christianson et al., 2006; Ferreira et al., 2002, 2001; Ferreira & Lowder, 2016) highlights the fact that the parser may perform superficial analysis of linguistic input based on heuristics, leading to inaccurate interpretations. For some extreme cases syntax seems to be completely bypassed (Traxler, 2011), while for other cases the interpretation derived from syntactic algorithm coexists with the interpretation derived from heuristics, resulting in a complex state of knowledge, a mixture of right and wrong interpretations. To illustrate this process, a series of experiments (Christianson et al., 2001, 2006; Ferreira et al., 2001) were conducted focusing on how readers recover from the garden-path effect when reading sentences such as "While Anna bathed the baby played in the crib." Theoretically speaking, the right interpretation for this sentence is that the baby played in the crib during the time when Anna bathed. However, before arriving at this final interpretation, readers may at first treat *baby* as the object of the verb *bathe* and later find this interpretation incompatible with the following verb phrase *played in the crib*, which forces the readers to adopt a different interpretation by taking *while Anna bathed* as the subordinate clause and *the baby played in the crib* as the main clause. An interesting finding in the above experiments was that the mental representation of the revised interpretation was not free from the previously experienced garden-path effect. When being asked whether or not the baby played in the crib, almost all participants correctly provided an affirmative answer. However, when the researchers probed participants' understanding of the subordinate clause, asking, for example, "did Anna bathe the baby?", participants inaccurately answered "yes" about 40% of the time (Ferreira & Lowder, 2016). Slattery et al. (2013) provided further evidence that the misinterpretations are not due to the failure in constructing the correct syntactic representation but the failure in replacing the earlier memory of incorrect interpretation with the correct interpretation derived later in time.

Similar mechanisms may also underlie the processing of nested epistemic expressions. When the second modal of the expression is processed, the epistemic strength of the first modal still lingers in memory, leading to a mixture of contrasting epistemic strengths which will eventually be normalized as an expression of uncertainty. Given that the epistemic strength of the two modals are mixed, the sequence in which two modals enters the processor does not matter too much. Thus, from the good-enough theory of sentence processing, it is possible that in everyday communication, interlocutors are not sensitive to the scope difference implied by the word order of the modals. In this way, interlocutors interpret the expression "He *certainly may* have forgotten" as having the same meaning as "He *may certainly* have forgotten." Thus, interlocutors would assign the same probability to the proposition "he has forgotten" regardless of whether they hear the utterance "He *certainly may* have forgotten" or "He *may certainly* have forgotten." That is to say, good-enough theory of sentence processing predicts that if we told someone "Tom *certainly may* have forgotten the meeting," and then asked them "How likely is it that Tom has forgotten the meeting?", they would provide the same probability rating as they were told "Tom *may certainly* have forgotten the meeting."

By contrast, the previously discussed scope account suggests that "Tom certainly may have forgotten the meeting" means *it is certainly the case that Tom may have forgotten the meeting* and the scope of *may* is embedded within the scope of the modal *certainly*. On the other hand, "Tom may certainly have forgotten the meeting" means *it may be the case that Tom certainly has forgotten the meeting*, and in this case, the scope of *certainly* is embedded within the scope of the modal *may*. This difference in meaning could be reflected from the probability rating of the embedded statement "Tom has forgotten the meeting" given nested epistemic expressions in different word orders. According to the well-established effect of anchoring heuristic (Epley & Gilovich, 2006; Jacowitz & Kahneman, 1995; Tversky & Kahneman, 1974), people's estimation of uncertainty is robustly influenced by the information that is initially provided, which "tends to exert drag on the subsequent adjustment process, leaving final estimates too close to the original anchor" (Epley & Gilovich, 2006, p. 311). A classic example of such effect is the difference in the estimation of an uncertain quantity when given different anchors (Jacowitz & Kahneman, 1995, p. 1163). When asking participants to estimate the number of bars in Berkeley, California, researchers found the estimate was much higher if the participants were provided with a high anchor at the beginning (e.g., "Is your estimate greater or smaller than 85? What is your estimate?") compared with the condition in which the participants were provided with a low anchor at the beginning (e.g. "Is your estimate greater or smaller than 10? What is your estimate?").

In terms of the judgment of nested epistemic expressions, if the scopes of the component modals are incrementally processed, the outer modal with a wider scope is processed before the inner modal, and thus its epistemic strength anchors the estimate of the epistemic strength of the inner modal. If such anchoring effect takes place during the processing of nonharmonic nested epistemic expressions such as "Tom *certainly may* have forgotten the meeting" and "Tom *may certainly* have forgotten the meeting," an interlocutor should perceive the version in which the modal with a higher epistemic strength comes first (which is *certainly may* in this example) as the one granting higher probability to the embedded statement ("Tom has forgotten the meeting.") compared with the nested expression in which the modal with a lower strength comes first (which is *may certainly* in this example). This is because when the modal with a higher epistemic strength comes first and thus scopes over the lower modal, the epistemic strength of this modal is processed before the epistemic strength of the inner modal and thus anchors the epistemic strength of the inner modal to a higher level. That is to say, the epistemic strength of the modal *may* is higher in the expression *certainly may* than in the cases where the modal *may* stands alone. On the other hand, when the modal with a lower epistemic strength comes first and thus scopes over the higher modal (such as *may certainly*), the epistemic strength of the lower modal is processed before the epistemic strength of the higher modal and thus anchors the epistemic strength of the inner modal to a lower value. In this case, the epistemic strength of the modal *certainly* is lower in the expression *may certainly* than in the cases where the modal *certainly* stands

alone. In short, if the scopes of the nested epistemic expressions have been thoroughly processed, the anchoring heuristic should boost the epistemic strength of the modal *may* in *certainly may* while keeping the epistemic strength of modal *certainly* unchanged; on the other hand, the same anchoring effect should lower the epistemic strength of the modal *certainly* in *may certainly*, while keeping the epistemic strength of *may* unchanged. As a result, *certainly may* should grant a higher probability to the embedded statement than *may certainly*.

In this study, we conducted a series of three experiments investigating how comprehenders process the semantics of nested epistemic expressions. Specifically, we focused on the modally nonharmonic combinations (such as *certainly may, must possibly*, and *probably might* etc.) asking whether or not the order of the two modals has an effect on comprehenders' interpretation of the expressions. Based on the anchoring effect we elaborated in the previous paragraph, we believe the epistemic strength as reflected from the probability ratings could be an informative proxy of how people process the scopes of the modals in a nested expression. If participants rate the epistemic strength of nested epistemic expressions differently based on the order of the two modals, especially if the nested expressions in which the modal with a higher epistemic strength comes first is rated higher than the nested expressions in which the modal with a lower strength comes first, it means that the scope relation of the two modals has been processed. However, if the order of the two nested modals does not influence how participants rate the epistemic strength of the expressions, then there is no evidence showing comprehenders process the scope relation of the two modals. In this case, the mechanism proposed by the good-enough processing framework is more likely to underlie the processing of nested epistemic expressions.

Experiment 1

In the first experiment, participants read English dialogues containing epistemic expressions, and for each of the dialogues they read, they were instructed to rate the probability of a statement based on the epistemic modal(s) presented in the dialogue. The order of the nested epistemic modals in the dialogues was manipulated to test whether readers interpret the probability of the embedded statement differently based on different ordering of the nested modals in the dialogues.

Participants

Fifty-two adult native English speakers (22 men and 30 women) were recruited from Amazon Mechanical Turk, a crowd-sourcing internet marketplace. All participants had a valid U.S. IP address and received monetary compensation for their participation in this study.

Stimuli

In this study, 16 experimental items were compiled into a questionnaire hosted by *Qualtrics*, an online survey platform. Each experimental item consisted of a written dialogue between two interlocutors followed by a question. The format of the dialogue was consistent across all experimental items. In each dialogue, the first speaker asked a question, and the second speaker provided an answer to the question. The words of the second speaker contained an epistemic expression. As shown in Figure 1, the second speaker's reply fell into four experimental conditions based on the way the epistemic expression was manipulated. In the *High-Low* condition, the epistemic modal expressing higher probability (which is *probably* 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 (*might probably*). The nested-modal conditions differed only in the word order of the two epistemic modals, and if participants formed

	Dialogue	Condition
Speaker 1	"Where is my blue shirt?"	
Speaker 2	"It probably might be in the bottom drawer of the dresser."	High-Low
	"It might probably be in the bottom drawer of the dresser."	Low-High
	"It might be in the bottom drawer of the dresser."	Low
	"It is probably in the bottom drawer of the dresser."	High
Question	How likely is it that the blue shirt is in the bottom drawer of the dresser?	

Figure 1. Experiment 1: Example of an experimental item.

different interpretations for sentences in *High-Low* and *Low-High* conditions, that would be an evidence that if more than one epistemic modals were presented in a clause, readers are sensitive to the scope of the modals during sentence processing. If participants formed identical interpretation for *High-Low* and *Low-High* condition, it would suggest that the scope of the modals is not thoroughly processed when comprehending nested epistemic expressions.

In addition to the nested-modal conditions, there were two single-modal conditions in which only one epistemic modal was present in the second speaker's words. In the *Low* condition, the sentence contained only the modal expressing lower probability (which is *might* in this example), and in the *High* condition, the sentence contained the modal expressing higher probability (which is *probably* 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. In this example, the nested epistemic modals were *might probably* (or *probably might*), in which the modal expressing higher probability was *probably*, while the modal expressing lower probability was *might*. Other experimental items may have a different combination of nested modals. In this study, we selected four pairs of nested modals, which were *must and probably, would and possibly, probably and might*, and *certainly and may*, and each pair of the modal combinations appeared in four different experimental items. These combinations each consisted of two epistemic modals, one expressing higher probability than the other, and all these combinations appeared in the *Corpus of Contemporary American English*, indicating their possible occurrence in American English.

Each experimental item ended with a question after the dialogue. The question probed participants' interpretation of what the second speaker said by asking the probability of a statement that had been mentioned in the dialogue. In this example, the question is "how likely is it that the blue shirt is in the bottom drawer of the dresser?", and the statement in question is "the blue shirt is in the bottom drawer of the dresser." Since the statement was embedded within the scope of the epistemic expressions, participants' judgments would be different depending on which version of the stimuli they saw. Participants were instructed to indicate the probability using a slider from zero, meaning "impossible," to 100, meaning "sure to happen." The inventory of all items in this experiment can be accessed via GitHub.[3]

There were 16 items in this experiment, and each experimental item appeared in one of four conditions. Four 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. During the experiment, all the items in the list were randomized.

Procedure

Participants used their own computer to access the link to 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 in the list was presented on the computer monitor screen. 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. The study ended after participants had answered the questions for all 16 experimental items.

Data analysis

The design of this experiment was treated as a single factor with four experimental conditions, and the probability ratings of items in different conditions were gathered and analyzed. Since the rating scores were bounded between zero and 100, we transformed the rating probability into its logit following the steps:

(1) Rescale the rating score from *0 to 100* into *0 to 1*
(2) Recode the rescaled variable, for which one is coded as 0.999, while zero is coded as 0.001, following Verkuilen and Smithson (2012)
(3) Perform logit transformation on each recoded rating score p using the equation

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

Bayesian mixed-effects models were constructed using R package brms (Bürkner, 2017) with default priors[4] to explore whether the order of the nested epistemic modals influences participants' interpretation of the expression. Those mixed-effects models included the condition of the epistemic expression (*High-Low, Low-High, Low*, and *High*) as the fixed effect, with both subjects and items as random effects, while the logit of probability ratings was treated as the dependent variable. The experimental conditions were dummy-coded, and in order to compare each nested-model condition with the single-model conditions, we set the reference level to the rating score of *High-Low* condition in one model (Table 1) and to *Low-High* condition in another model (Table 2), following the same model structure:

$$rating\ logit = 1 + condition + (1 + condition|subject) + (1 + condition|item)$$

The data and script used for statistical analysis of this study are available in GitHub.[5]

Table 1. Experiment 1: Estimate of intercept and slopes from Bayesian mixed-effects model with High-Low condition as the baseline.

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.87	0.18	0.53	1.23
Low-High	−0.20	0.16	−0.52	0.12
High	1.18	0.38	0.45	1.93
Low	−0.53	0.19	−0.91	−0.15

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

Table 2. Experiment 1: Estimate of intercept and slopes from Bayesian mixed-effects model with Low-High condition as the baseline.

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.68	0.11	0.46	0.91
High-Low	0.20	0.21	−0.22	0.62
High	1.38	0.42	0.55	2.20
Low	−0.33	0.14	−0.59	−0.06

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

Results

When participants read stimuli in which only one single epistemic modal was presented, the probability rating of the statement reflected participants' knowledge about the epistemic strength of the modal. Linguistic theories (Halliday & Matthiessen, 2004; Holmes, 1982) suggested a three-point scale of epistemic strength from maximal certainty (such as the word *certainly*) to medium certainty (such as *probably*) to minimum certainty (such as *may*). As can be seen from Figure 2, participants' understanding of the degree of certainty expressed by various epistemic modals was largely consistent with what previous theories described. When the word *possibly, may*, and *might* appeared in the dialogue, participants rated the statement with least certainty. On the other hand, when the word *certainly, would*, and *must* were used in the dialogue, participants rated the statement with the highest certainty. Moreover, all rating scores in this study were above 50 out of 100, meaning that the use of the above epistemic modals made the embedded statements sound more likely than the chance level.

A noticeable discrepancy between participants' rating and the prediction of linguistic theory is that whereas Halliday and Matthiessen (2004) proposed that the word *would* expressed medium certainly, similar to the word *probably*, participants interpreted "would *S*" (*S* for a statement) as "it is almost certain that ," rather than "it is probably the case that *S*." Taking into consideration that Halliday and Matthiessen (2004) was not based on American English and that 20 years have passed since its publication, it is not completely unexpected that the interpretation of some epistemic modals is now slightly different from what was believed 20 years ago. Since this study focused on the processing of nested epistemic expressions, the experimental manipulation of *High-Low* and *Low-High* conditions would make sense as long as the modals within a pair of combination did not express the same degree of certainty. The single-modal ratings showed that for all modal combinations we had in this experiment, one modal indeed expressed higher probability than the other modal. To be more specific, modals described as expressing high degree of probability (*certainly* and *must*) on average were rated 7

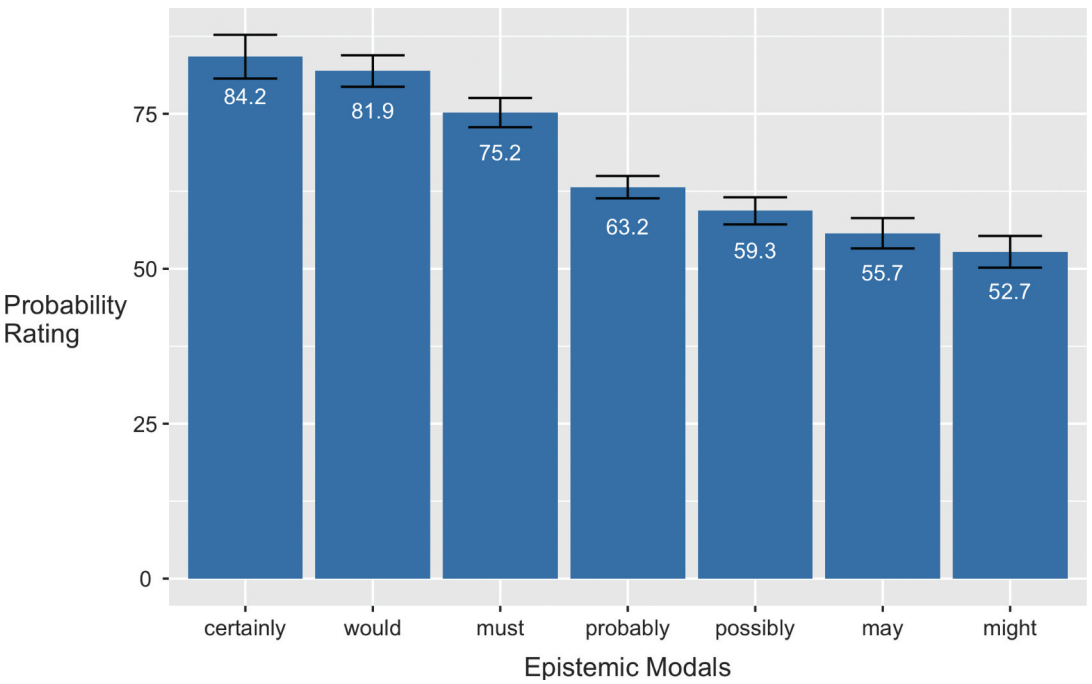

Figure 2. Experiment 1: Mean probability rating of different epistemic modals (with standard error).

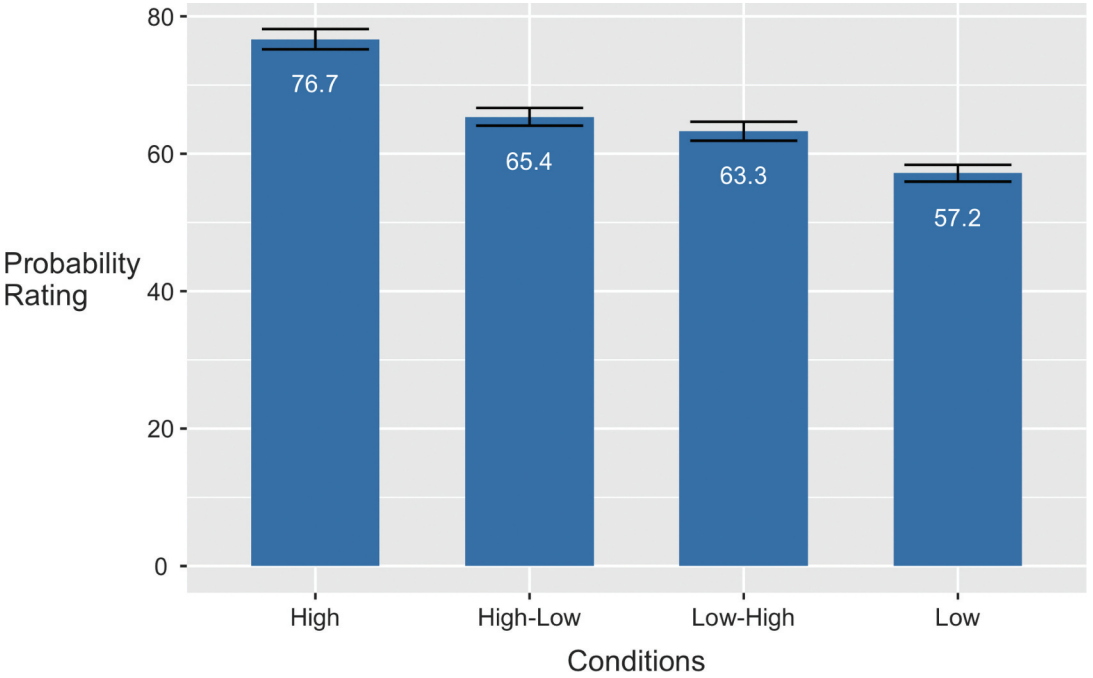

Figure 3. Experiment 1: Mean probability rating of different experimental conditions (with standard error).

points higher than modals that were believed to express medium degree of probability (*would* and *probably*), which were rated 16 points higher than modals expressing low degree of probability (*possibly, may*, and *might*).

The probability rating for items in different experimental conditions is shown in Figure 3. When averaging across all experimental items, we found that participants gave the highest rating of probability to items containing a single modal of greater epistemic strength. For this experiment, the modals in *High* conditions were *certainly, must, would*, and *probably* (when paired with the modal *might*), and the average probability rating for them was about 76.7 out of 100. On the other hand, participants gave the lowest rating of probability to items containing a single modal of less epistemic strength. For this experiment, the modals in *Low* conditions were *possibly, might, may*, and *probably* (when paired with the modal *must*). The mean probability rating score for items containing these modals were 57.2. The rating score of the nested-modal conditions lay in between the rating scores of single-modal conditions, with *High-Low* condition (65.4) rated two points higher than the *Low-High* condition (63.3).

A closer look at the two nested-modal conditions revealed that the small difference between the *High-Low* and *Low-High* condition in probability rating was not statistically meaningful. Two Bayesian mixed-effects models were constructed estimating the logit of the probability rating score across conditions following the steps illustrated in the data analysis section . When *High-Low* condition was the baseline for comparison, participants' rating of the baseline was statistically lower than that of the *High* condition ($\beta = 1.18$, 95% CI = [0.45, 1.93]) while higher than that of the *Low* condition ($\beta = -0.53$, 95% CI = [−0.91, −0.15]). However, the 95% credible interval for the slope of *Low-High* contained the value of zero ($\beta = -0.2$, 95% CI = [−0.52, 0.12]), meaning it is likely that there was no statistical difference between the *High-Low* condition and *Low-High* condition (Table 1). The same pattern was observed when the *Low-High* condition was made the baseline for comparison

(Table 2). The probability rating for both the *High* condition (β = 1.38, 95% CI = [0.55, 2.2]) and the *Low* condition (β = −0.33, 95% CI = [−0.59, −0.06]) was statistically different from the baseline but not the *High-Low* condition (β = 0.20, 95% CI = [−0.22, 0.62]).

Following the suggestion from an anonymous reviewer, Bayes factor analyses were conducted to further evaluate the odds that there was no difference between the *High-Low* and *Low-High* conditions. We specifically focused on the coefficient of the *Low-High* condition which indicated how much the probability rating of the *Low-High* condition differed from that of the baseline (*High-Low*) condition. The null hypothesis (H0) was that the coefficient of the *Low-High* condition equaled zero, while the alternative hypothesis (H1) stated that the coefficient was not zero. The Bayes factor in favor of the H0 (BF01) was calculated with the R package brms (Bürkner, 2017) using the Savage-Dickey density ratio method (Wagenmakers et al., 2010). When calculating BF01, we selected priors from six normal distributions with a mean of zero and minus 0.2 (the estimate of the coefficient from the previous Bayesian mixed-effects models (Table 1)) and a standard deviation of 10, 5, and 1, respectively. Bayes factor analyses using the priors specified above revealed moderate to strong evidence for the null hypothesis (Table 3). Since Bayes factor analyses are believed to be sensitive to prior selections (Nicenboim & Vasishth, 2016), we also explored the effect of more informative priors on BF01 and found that the evidence in support of the null hypothesis was not overly influenced by the prior specification.[6]

Discussion

The findings of this experiment showed that participants were sensitive to the epistemic strength of epistemic modals in the dialogue, and thus they made reasonable inferences about the probability of a statement embedded within the scope of a single epistemic modal. For example, the sentence "He may be in the candy shop" and "He is certainly in the candy shop" both indicate the probability of *he was in the candy shop*. This means the same proposition "he is in the candy shop" is embedded within the scope of the epistemic modal in both sentences. The difference between these two sentences lies in the semantics of the modal. While *may* expresses low probability, the word *certainly* expresses maximal 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. Thus in this experiment, when there was only one epistemic modal in the dialogue, the probability rating of the statement decreased if it was embedded within the word *may*, compared to being embedded within the word *certainly*.

When there were two epistemic modals in a sentence, the probability of the embedded statement lay in between the epistemic strength of the component modals. If the sentence "He may be in the candy shop" expressed low probability, while "He is certainly in the candy shop" expressed high probability (Halliday & Matthiessen, 2004, p. 623), then both "He may certainly be in the candy shop" and "He certainly may be in the candy shop" expressed medium degree of probability.

Table 3. Experiment 1: Summary of the priors, estimates and Bayes factors in favor of the null hypothesis.

Prior	Estimate	BF01
Normal (0, 10)	−0.20	27.96
Normal (0, 5)	−0.20	15.01
Normal (0, 1)	−0.19	3.08
Normal (−0.2, 10)	−0.20	30.01
Normal (−0.2, 5)	−0.19	14.75
Normal (−0.2, 1)	−0.20	3.11

[a]Based on Lee and Wagenmakers (2014, p. 105), a BF01 of 3–10 is considered as a moderate evidence for H0, while a BF01 of 10–30 is considered as a strong evidence for H0.

However, there was no difference between *certainly may* and *may certainly* in terms of the probability expressed. 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, which was revealed from the credible interval of the coefficients (see Tables 1 and 2 for the coefficients of the nested-modal conditions) and further confirmed by Bayes factor analyses. Since the scope account of nested epistemic modals predicts a difference in probability rating between the two nested-modal conditions, we did not find supporting evidence for the scope account of the nested epistemic expressions in this experiment.

It is possible that during language processing, interlocutors treat nested epistemic expressions as an indicator of medium degree of uncertainty. In this experiment, the average rating score of nested-modal conditions was about 64 out of 100 (Figure 3), similar to the degree of probability expressed by the word *probably* (Figure 2). From this perspective, the sentence "He may certainly be in the candy shop" and "He certainly may be in the candy shop" both mean something similar to "He is probably in the candy shop." Before arriving at such conclusion, we need to carefully consider the possibility that there is a difference between the *High-Low* and *Low-High* condition and that somehow the experimental stimuli we used failed to elicit the order effect.

Specifically, it might be the case that for the modals we selected in this experiment, the *higher* modals and *lower* modals were not so different in terms of their epistemic strength, and thus the first modal failed to be a strong anchor influencing the interpretation of the second modal. As mentioned above, the distance between categories on the three-point scale of epistemic strength is not evenly divided. Modals described in previous literature as expressing high degree of probability (*certainly* and *must*) on average were rated 7 points higher than modals that were believed to express medium degree of probability (*would* and *probably*), while the modals expressing medium degree of probability were rated 16 points higher than modals expressing low degree of probability (*possibly, may*, and *might*). As a result, if in a nested epistemic expression the higher modal was from the high extreme on the scale while the lower modal was from the medium category on the scale, the difference in epistemic strength between these two modals might not be obvious. This was the case for one of the modal combinations used in this experiment (*must and probably*). According to Halliday and Matthiessen (2004), *must* expresses high probability while *probably* expresses medium probability. They belong to adjacent categories in the three-point scale of epistemic strength. If the two modals in the nested-conditions were all selected from the opposite extremes in the scale, such as *certainly* (which belongs to the high extreme) and *might* (which belongs to the low extreme), the anchoring effect of the first modal would be much stronger, and thus, the change in word order might be able to change participants' interpretation of the nested epistemic expression. This hypothesis was tested in the second experiment.

Experiment 2

This experiment tested the hypothesis that the order of the two modals affects interlocutors' interpretation of nested epistemic expressions if one modal expresses very high probability while the other modal expresses very low probability. The research paradigm used in this experiment was largely the same as that of the first experiment with the crucial difference that for this experiment, the two modals in a pair were selected from the high and low extremes on the scale of probability. Thus, any statistically meaningful difference in rating score between nested modals of the opposite word order would serve as an evidence that word order affects the interpretation of nested epistemic expressions. The scope account further predicts that the nested expressions in the *High-Low* word order would have a higher rating than nested expressions in the *Low-High* order.

Participants

Sixty college students (44 women, 16 men) 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.

Stimuli

This study contained 16 experimental items, each of which was a written dialogue between two interlocutors, followed by a question. The structure of the dialogue and the manipulation of the experimental conditions were identical to those of the first experiment, though the topics of the dialogues were not the same. The crucial difference between this experiment and the first experiment lay in the modal combinations contained in the dialogue. In this study, we selected four pairs of nested modals, which were *definitely and may, definitely and might, certainly and might, certainly and may*. Each of these combinations consisted of two epistemic modals, one expressing very high probability while the other expressing very low probability. All of these modal combinations had been found in *Twitter* or *Corpus of Contemporary American English*, indicating their possible occurrence in American English. Each of the modal combinations appeared in four different experimental items, and an example of the experimental item is represented in Figure 4.

Four lists of experimental items were created following the same Latin square design as the first experiment to ensure that each list contained an equal number of items in each condition, while each experimental item only occurred once in a list. In this experiment, each list also included 48 filler items to hide the intended research question from the participants. Similar to the experimental items, each filler item was also a dialogue followed by a possibility judgment question. However, the filler item did not contain any epistemic modals or it only contained the negation of an epistemic modal (such as *impossible* or *might not*). All four lists contained the same set of filler items, and by adding the experimental items there were 64 items in total for each list. During the experiment, all items in a list were randomized. The complete list of items used in this study can be found in the same *GitHub* inventory listed above.

Procedure

Participants were tested in person in the research lab and were seated in front of a desk computer in a testing room. The web page hosting this study was presented on the monitor of the desk computer. Participants completed the study following the same steps as those of the first experiment.

	Dialogue	Condition
Speaker 1	"The house smells very bad. I wonder if something is rotting here"	
Speaker 2	"There certainly may be something rotting in the house."	High-Low
	"There may certainly be something rotting in the house."	Low-High
	"There may be something rotting in the house."	Low
	"There certainly is something rotting in the house."	High
Question	How likely is it that something is rotting in the house?	

Figure 4. Experiment 2: Example of an experimental item.

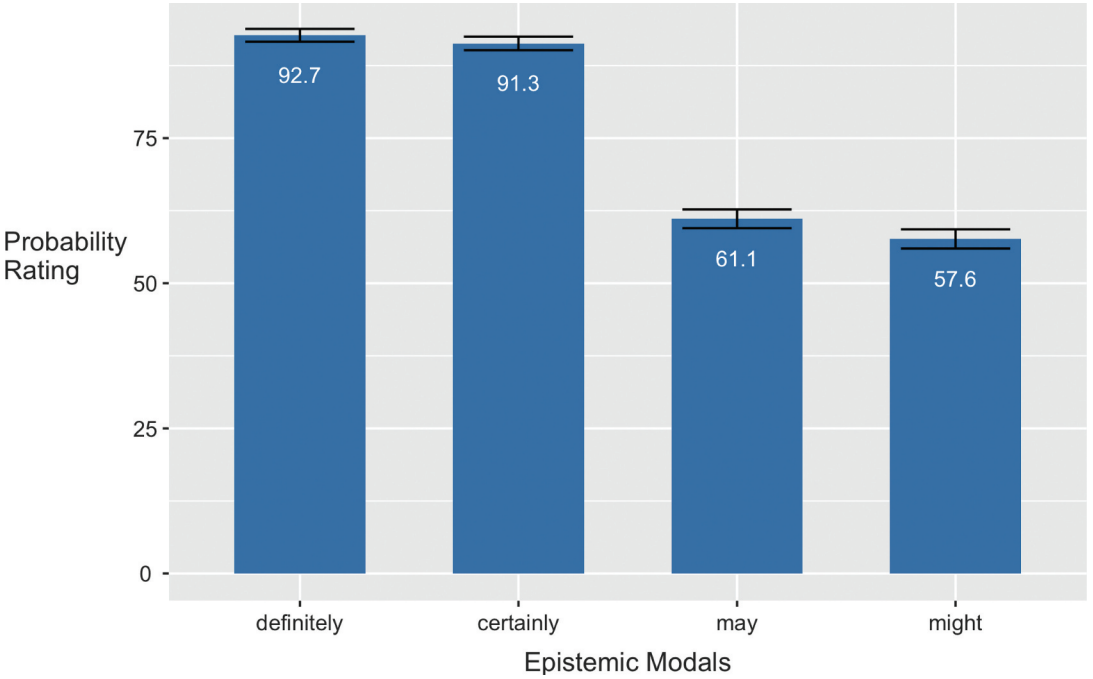

Figure 5. Experiment 2: Mean probability rating of different epistemic modals (with standard error).

Data analysis

This study followed the same data analysis procedure as the first experiment.

Results

The probability rating scores of individual modals in this study are shown in Figure 5. As can be seen from the figure, when a statement was embedded within the scope of *may* or *might*, participants rated the probability of that statement as about 60 out of 100. By comparison, when a statement was embedded in the scope of *definitely* or *certainly*, participants rated the probability of that statement as about 90 out of 100. In this study, the modal of *may* and *might* belonged to the *Low* condition, while the modal of *definitely* and *certainly* belonged to the *High* condition. The difference in epistemic strength between *High* modals and *Low* modals was apparent.

This pattern was also reflected from the average rating scores across items in different conditions (Figure 6). We found that participants gave the highest rating of probability for the experimental items that contained a single modal of high epistemic strength (92 out of 100). On the other hand, participants gave the lowest rating of probability for items containing a single modal of low epistemic strength (59.4 out of 100). The difference in rating score between *High* and *Low* condition in this experiment was 32.6, which was much greater than the difference between the two conditions in the first experiment (which was 19.5).

The rating score of the nested-modal conditions (Figure 6) lay in between the rating score of single-modal conditions, with *High-Low* condition (70.5) rated one point higher than the *Low-High* condition (69). However, this small difference was not statistically meaningful. Following the same procedure as the first experiment, we constructed two Bayesian mixed-effects models to estimate the logit of the probability rating score across conditions. When the *High-Low* condition was the baseline for comparison, participants' rating of the baseline was statistically lower than that of the *High* condition (β = 2.70, 95% CI =

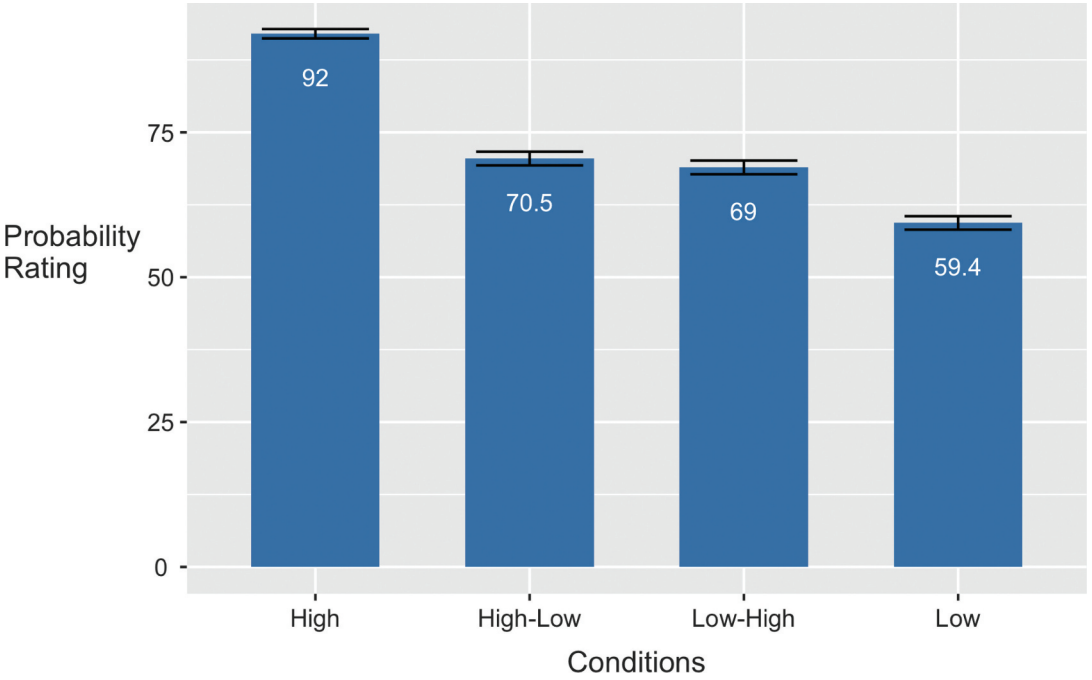

Figure 6. Experiment 2: Mean probability rating of different experimental conditions (with standard error).

Table 4. Experiment 2: Estimate of intercept and slopes from Bayesian mixed-effects model with High-Low condition as the baseline.

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.19	0.15	0.89	1.49
Low-High	−0.12	0.12	−0.37	0.12
High	2.70	0.22	2.26	3.14
Low	−0.66	0.13	−0.91	−0.40

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

Table 5. Experiment 2: Estimate of intercept and slopes from Bayesian mixed-effects model with Low-High condition as the baseline.

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.06	0.15	0.78	1.34
High-Low	0.12	0.14	−0.14	0.40
High	2.83	0.23	2.39	3.28
Low	−0.53	0.12	−0.76	−0.31

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

[2.26, 3.14]) while higher than that of the *Low* condition (β = −0.66, 95% CI = [−0.91, −0.40]). However, the 95% credible interval for the slope of *Low-High* contained the value of zero (β = −0.12, 95% CI = [−0.37, 0.12]), meaning it is likely that there was no statistical difference between the *High-Low* condition and *Low-High* condition (Table 4). The same pattern was observed when the *Low-High* condition was made the baseline for comparison (Table 5). The probability rating for both *High* condition (β = 2.83, 95% CI = [2.39, 3.28]) and *Low* condition (β = −0.53, 95% CI = [−0.76, −0.31]) was statistically different from the baseline but not the *High-Low* condition (β = 0.12, 95% CI = [−0.14, 0.40]).

Bayes factor analyses were conducted to further evaluate the odds of no difference between the *High-Low* and *Low-High* conditions. We focused on the coefficient of the *Low-High* condition which indicated how much the probability rating of the *Low-High* condition differed from that of the baseline (*High-Low*) condition. The null hypothesis (H0) was that the coefficient of the *Low-High* condition equaled zero, while the alternative hypothesis (H1) stated that the coefficient was not zero. The Bayes factor in favor of the H0 (BF01) was calculate following a procedure similar to what has been reported in the Results section of Experiment 1, with the only difference being the choice of priors. We selected six priors from six normal distributions with a mean of zero and minus 0.12 (the estimate of the coefficient from the previous Bayesian mixed-effects models; Table 4), respectively, and a standard deviation of 10, 5, and 1, respectively. Bayes factor analyses using the priors specified above revealed moderate to strong evidence for the null hypothesis (Table 6). Sensitivity analyses with more informative priors confirmed that the evidence in support of the null hypothesis was not overly influenced by the prior specification.[7]

Discussion

The aim of this experiment was to test the hypothesis that when the two modals in a nested epistemic expression are robustly different in their individual epistemic strength, readers will start to process the scope of each modal, assigning different interpretations to the nested expressions which contain the same modals but of the opposite word order. For each of the modal pairs we used in this study, one modal expressed very high probability and the other modal expressed very low probability. We believed that such a contrast in epistemic strength would boost the anchoring effect of the first modal and thus make the order of the two modals more salient to participants. To be more specific, according to Epley and Gilovich (2006), the processing of the initial information drags the subsequent adjustment process. When the modal expressing a very high probability comes first, it drags the probability expressed by the second modal more toward the high end of the scale of epistemic strength; on the other hand, when the modal expressing a very low probability comes first, it drags the interpretation of the second modal more toward the low end of the scale of epistemic strength. This should result in a bigger semantic difference between the *High-Low* and *Low-High* condition.

The results of the experiment showed that participants were able to detect the increased difference in epistemic strength between the two modals in a nested expression. They gave a higher rating score for modals in the *High* condition (92 out of 100) compared with that of the first experiment (76.7 out of 100), and as a result, the difference between the *High* and *Low* condition in this experiment was much greater than the difference in the first experiment (32.6 vs 19.5). The mean rating scores for the two nested conditions in this experiment (70.5 and 69) were also slightly higher than those of the first experiment (65.4 and 63.3). However, in the second experiment, the patterns we found across experimental conditions were essentially the same as the

Table 6. Experiment 2: Summary of the priors, estimates and Bayes factors in favor of the null hypothesis.

Prior	Estimate	BF01
Normal (0, 10)	−0.12	49.95
Normal (0, 5)	−0.12	24.32
Normal (0, 1)	−0.12	5.23
Normal (−0.12, 10)	−0.12	47.64
Normal (−0.12, 5)	−0.12	23.91
Normal (−0.12, 1)	−0.12	5.07

[a]Based on Lee and Wagenmakers (2014, p. 105), a BF01 of 3–10 is considered as a moderate evidence for H0; a BF01 of 1030 is considered as a strong evidence for H0, while a BF01 of 30–100 is considered as a very strong evidence for H0.

patterns revealed in the first experiment. The rating scores of the two nested conditions lay in between the ratings of *High* and *Low* conditions, while no statistical difference was found between the rating scores of the two nested conditions.

It seems that when readers saw a statement within the scopes of two epistemic modals, one expressing high probability and the other expressing low probability, the readers would assign to the embedded statement a medium probability in between the epistemic strength of the individual modals. The order of the two modals did not affect the interpretation of the embedded statement, which failed to support the scope account which suggested that meaning of the expression will change if the order of the modals changes (Lyons, 1977; Moss, 2015). The patterns observed in the previous experiments were more consistent with the good-enough processing account, which claimed that the processor does not always compute every piece of information in the language input following a rigid parsing algorithm; rather, the processor may form superficial interpretation of the input that is not completely faithful to the linguistic representations that are assumed to underlie the forms (Ferreira et al., 2002; Ferreira & Lowder, 2016).

It is possible that in the previous two experiments, readers treated the two modals in a nested expression as a single lexical unit of idiomatic meaning. A similar case would be the use of double modals *might could* in some southern dialects of American English (Di Paolo, 1989). If this is true, as long as the two modals are adjacent to each other, no matter how different they are in epistemic strength, the change of the word order would not change the meaning of the expression. However, it is possible that when the distance of the two modals is enlarged, the word order starts to matter. One way of enlarging the distance between the modals is inserting a parenthetical element in between the two modals. Compare the sentence "Bob *definitely might* have hit traffic on his way home" with "Bob *definitely*, according to the radio, *might* have hit traffic on his way home." In the second sentence, the parenthetical element *according to the radio* pulls the two modals apart. It is possible that, in this case, readers will process the scope of the two modals and interpret the sentence differently depending on which modal they see first. The third experiment of this study investigated this possibility.

Experiment 3

This experiment tested the hypothesis that the order of the two modals affects interlocutors' interpretation of nested epistemic expressions if the two modals in question are not adjacent to each other but separated by other words in between. The research paradigm used in this experiment was the same as that of the first two experiments, and the crucial difference was that for this experiment the two modals in a pair were separated by a parenthetical element. The inclusion of parenthetical elements inhibited the parser from treating the two epistemic modals as a single lexical item and also provided the parser with more time to process the meaning of the first modal before encountering the second modal, thus the anchoring effect of the first modal would be more salient. We believed that with this adjustment, the scope of the two modals in the stimuli would become more salient to the participants. Any statistically meaningful difference in rating score between nested modals of the opposite word order would serve as an evidence that the scopes of the modals are processed.

Participants

Sixty-one college students (43 women, 18 men) 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.

Stimuli

This study contained 32 experimental items, each of which consisted of a written dialogue and a probability judgment question. The structure of the dialogue and the manipulation of the experimental conditions were identical to those of the first two experiments. The major difference was that in this study the two modals in the dialogue were separated by a parenthetical element indicating the source of information that the second speaker relied on when answering the first speaker's question, such as "according to the weather forecast," "based on my experience," and so on. For nested-modal conditions these words concerning the information source appeared in-between the two modals, while for the single-modal conditions these words appeared at the beginning of the sentences as in "According to the weather forecast, it would be windy." Participants were asked to rate the probability of a statement based on their interpretation of the epistemic expression in the dialogue.

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 of these combinations consisted of two epistemic modals, one expressing higher probability and the other expressing lower probability. All these modal combinations had been found in *Twitter* or *Corpus of Contemporary American English*, and each of the modal combination appeared in four different experimental items. An example of the experimental item is represented in Figure 7.

Four lists of experimental items were created following the same Latin square design as the previous two experiments. In addition to experimental items, each list contained 32 filler items, which were dialogues without epistemic auxiliaries or adverbs, while having parenthetical phrases that mimicked the structure of experimental items. All four lists contained the same set of filler items, with 64 items in total for each list. During the experiment, all items in a list were randomized. The complete list of items used in this study can be found in the same *GitHub* inventory listed above.

Procedure

Participants completed the study following the same procedure as the second experiment.

Data analysis

This study followed the same data analysis procedure as the first two experiments.

	Dialogue	Condition
Speaker 1	"Bob hasn't arrived yet. Do you think he hit traffic on his way home?"	
Speaker 2	"There definitely, according to the radio, might be a traffic jam on Bob's way home."	High-Low
	"There might, according to the radio, definitely be a traffic jam on Bob's way home."	Low-High
	"According to the radio, there might be a traffic jam on Bob's way home."	Low
	"According to the radio, there is definitely a traffic jam on Bob's way home."	High
Question	How likely is it that there is a traffic jam on Bob's way home?	

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

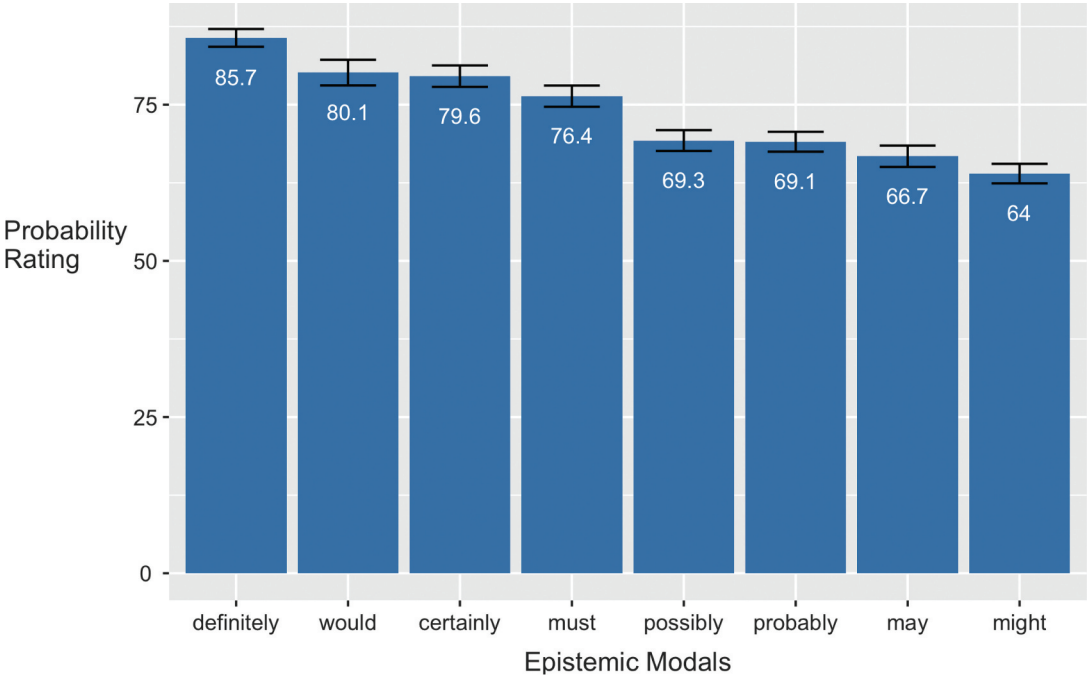

Figure 8. Experiment 3: Mean probability rating of different epistemic modals (with standard error).

Results

The summary of probability rating scores of individual modals in this study is shown in Figure 8. Similar to the previous two experiments, the rating scores of individual modals revealed a scale from high probability to low probability, which corresponded to participants' knowledge of each modal's epistemic strength. A closer inspection of the scale showed two noticeable differences from the first experiment. The first difference was the epistemic strength of the modal *probably* and *possibly*. In the first experiment, statements embedded in the scope of *probably* and *possibly* received probability ratings of 63.2 and 59.3, respectively, while in this experiment the rating scores of *probably* and *possibly* were almost identical to each other (69.1 and 69.3, respectively).

We constructed a Bayesian mixed-effects model to explore whether or not there was a statistical difference in probability rating score (in the logit scale) between items containing a single modal of *probably* and items containing a single modal of *possibly*. It turned out that the difference between the rating of *probably* and *possibly* observed in the first experiment was not statistically meaningful ($\beta = 0.14$, 95% CI = [−0.37, 0.66]). In this sense the pattern pertaining to the rating of *probably* and *possibly* observed in this experiment was essentially the same as the pattern observed in the first experiment.

	Modal: May	Modal: Might
Experiment 1	"He may have forgotten"	"He might be Mr. Gomes"
Experiment 3	"Based on my experience, the soup may have some star anise in it"	"According to her friends, she might have gotten her hair cut from the new salon downtown"

Figure 9. A comparison between two stimuli in Experiment 1 and Experiment 3 pertaining to the modal of *may* and *might*.

The second difference is that in this study, the modals of low epistemic strength, such as *may* and *might*, received higher rating scores (about 65 out of 100) compared to the scores in the first experiment (about 55 out of 100). This difference is largely due to the experimental manipulation, which included in-between modals a parenthetical element indicating information source. For single modal conditions, the same information source was mentioned at the beginning of the sentence preceding the epistemic modal. Figure 9 shows an example of stimuli containing modal *may* and *might* in the first and the third experiment. As can be seen from this example, items in the third experiment contained adverbial phrases that preceded the epistemic modals, indicating the source of information that the interlocutor relied on when evaluating the probability of the event, such as "based on my experience" or "according to her friends." By contrast, the stimuli of the first experiment did not include any indication of the information source, and thus when participants read words such as "He may have forgotten" or "He might be Mr. Gomes", they were not sure about the basis of these statements. The result of the third experiment suggested that providing the source of information boosted participants' estimation of the probability, especially for low modals.

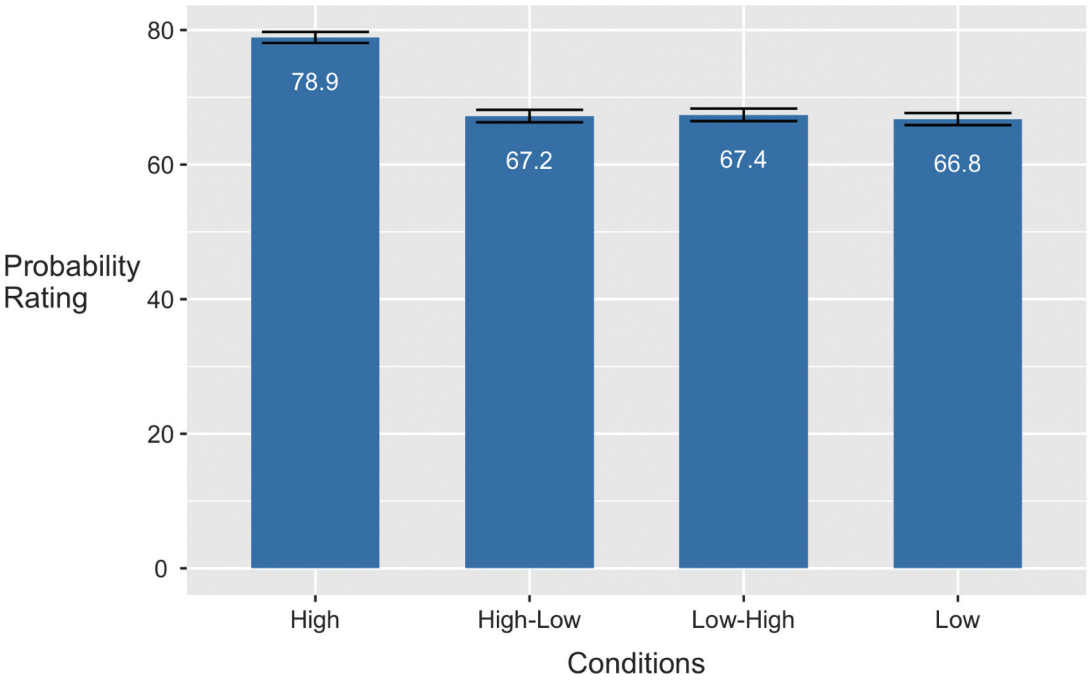

Figure 10. Experiment 3: Mean probability rating of different experimental conditions (with standard error).

Table 7. Experiment 3: Estimate of intercept and slopes from Bayesian mixed-effects model with High-Low condition as the baseline.

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.91	0.13	0.66	1.17
Low-High	0.01	0.08	−0.14	0.17
High	0.98	0.15	0.69	1.26
Low	−0.06	0.08	−0.21	0.09

[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 from Bayesian mixed-effects model with Low-High condition as the baseline.

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.93	0.13	0.68	1.18
High-Low	−0.01	0.08	−0.16	0.14
High	0.97	0.14	0.69	1.25
Low	−0.07	0.08	−0.22	0.08

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

The inflated rating scores of the low modals can also be observed across experimental conditions (Figure 10). While participants gave the highest rating of probability for the experimental items that contained a single modal of high epistemic strength (78.9 out of 100), they rated statement containing a single low modal (66.8) and nested modals (67.2 and 67.4) as equally likely. By comparison, for both the first and second experiments, the probability rating scores of the single low modal condition were below 60, significantly lower than the rating scores of the nested modal conditions.

Following the same procedure as the first two experiments, we constructed Bayesian mixed-effects models to estimate the logit of the probability rating scores across conditions. When *High-Low* condition was the baseline for comparison, participants' rating of the baseline was statistically lower than that of the *High* condition ($\beta = 0.98$, 95% CI = [0.69, 1.26]). However, there is no statistical difference between the baseline and the *Low* condition ($\beta = -0.06$, 95% CI = [−0.21, 0.09]) or in between the baseline and the *Low-High* condition ($\beta = 0.01$, 95% CI = [−0.14, 0.17]) (Table 7). The same pattern was observed when the *Low-High* condition was made the baseline for comparison. The probability rating for the *High* condition ($\beta = 0.97$, 95% CI = [0.69, 1.25]) was statistically different from that of the baseline; however, the rating scores of the *High-Low* condition ($\beta = -0.01$, 95% CI = [−0.16, 0.14]) and the *Low* condition ($\beta = -0.07$, 95% CI = [−0.22, 0.08]) were essentially the same as the baseline (Table 8).

To further evaluate the odds that there was no difference between the probability rating of the two nested conditions, Bayes factor analyses were conducted following a procedure similar to what has been reported in the first two experiments, with the only difference being the choice of priors. We selected six priors from six normal distributions with a mean of zero and 0.01 (the estimate of the coefficient from the previous Bayesian mixed-effects models; Table 7), respectively, and a standard deviation of 10, 5, and 1, respectively. Bayes factor analyses using the priors specified above revealed strong evidence for the null hypothesis (Table 9). Sensitivity analyses with more informative priors confirmed that the evidence in support of the null hypothesis was not overly influenced by the prior specification.[8]

Table 9. Experiment 3: Summary of the priors, estimates and Bayes factors in favor of the null hypothesis.

Prior	Estimate	BF01
Normal (0, 10)	0.01	127.11
Normal (0, 5)	0.01	61.73
Normal (0, 1)	0.01	13.05
Normal (0.01, 10)	0.01	134.08
Normal (0.01, 5)	0.01	68.62
Normal (0.01, 1)	0.01	13.26

[a]Based on Lee and Wagenmakers (2014, p. 105), a BF01 of 10–30 is considered as a strong evidence for H0; a BF01 of 30,100 is considered as a very strong evidence for H0, while a BF01 bigger than 100 is considered as an extreme evidence for H0.

Discussion

In this experiment, we tested the hypothesis that when the two modals in a nested epistemic expression are not adjacent to each other, readers would start to process the scope of each modal, assigning different interpretations to the nested expressions containing the same modals but in the opposite word order. For each of the modal pairs in this study, we inserted a parenthetical element in between the two modals to test whether when the distance between the two modals was enlarged, the order of the two modals would be more salient to participants. This manipulation addressed one concern we had for the previous experiments, which was that the two modals in a nested expression may be treated as a single idiomatic expression if they are adjacent to each other. A similar case is the use of double modals like *might could* in northern England and southern United States (Nagle, 2012). Among various linguistic analyses attempting to account for the structure of double modals (Battistella, 1995; Di Paolo, 1989; Elsman & Dubinsky, 2009), Di Paolo (1989) argued that the double modals is one single lexical item consisting of two words, similar to a compound.

The nested epistemic expressions investigated in the first two experiments are different from the double modals like *might could* in certain important aspects. In terms of syntactic categories, one modal in the nested epistemic expression is a modal auxiliary and the other is a modal adverb. As to the double modals, both of them are modal auxiliaries. In terms of semantic categories, the nested epistemic expression consists of two epistemic modals, while double modals consist one epistemic modal and one nonepistemic (such as deontic) modal (Nagle, 1994). Although there are noticeable differences between these two linguistic constructions, it is still possible that in the first two experiments, readers treated the two modals in a nested expression as a single lexical unit of idiomatic meaning. By inserting a parenthetical element in between the two modals, we increased the distance between them so that the parser would not treat the two modals as a single lexical item. Moreover, the increased distance between two modals provides the parser with more time to process the meaning of the first modal before encountering the second modal. Thus, the scope of the two modals becomes more salient. However, similar to the previous two experiments, there was still no statistical difference in rating score between the two nested-conditions in this experiment. When reading a statement containing more than one epistemic modals, participants rated that statement as less probable than the statement containing a single epistemic modal expressing high probability. The order of the two modals in the nested expression didn't affect the probability rating of the statement, even when the two modals were not adjacent to each other.

Similar to the previous two experiments, the pattern we observed in this experiment did not support the scope account of the nested epistemic expression. At least, participants were not sensitive to the supposed scope difference between the first modal and the second modal during the processing of nested epistemic expressions. The explanation we offered for the previous experiments was that the two modals were treated together as a single lexical unit expressing medium probability. However, for this experiment, the two modals in the stimuli were separated by parenthetical elements, and the enlarged distance between them made it less likely for participants to treat the two modals as one idiomatic expression. A more plausible explanation would be that during the processing of nested epistemic expressions, the epistemic strength of the first modal lingered in memory and was mixed with the epistemic strength of the second modal when it was encountered, a cognitive mechanism similar to the lingering misinterpretation of competing syntactic representation (Christianson et al., 2001, 2006; Ferreira et al., 2001; Slattery et al., 2013). The parser then treated the mixed epistemic strength as an indicator of general uncertainty without further pinning down the scope of each modal.

In this experiment, the purpose of including parenthetical elements was to increase the distance of the two modals; however, those parenthetical elements were not neutral in terms of the probability they implied. Parenthetical elements used in this study were expressions indicating the source of information on which the second speaker's statement was based. According to linguistic theories,

these parenthetical elements belong to the notional category of *evidentials*, expressions or grammatical markers that "indicate something about the source of the information in the proposition" (Bybee, 1985, p. 184). Evidentials are often grammaticalized in other languages as inflectional morphemes (Chafe & Nichols, 1986), and it has been estimated that about one-fourth of world's languages have grammatical evidentiality (Aikhenvald, 2004), especially for languages in North and South America, Caucasian languages, and Tibeto-Burman languages (Song, 2018). For some languages, the marking of information source using an inflectional morpheme is obligatory in statements (Bybee, 1985), while in English, evidentials are not grammaticalized as a part of morphological system, and speakers express the notion of evidentiality using words and phrases (De Haan, 2001; Gisborne & Holmes, 2007).

Admittedly, the inclusion of the information source in the dialogue could influence readers' rating of probability. However, since the parenthetical element was held constant across experimental conditions of the same experimental item, we were able to derive the relative rating differences across experimental conditions. Given there was still no difference in the rating scores between the two nested-modal conditions, we are confident about the consistent patterns observed across all three experiments reported in this study: While participants clearly differentiated the epistemic strength of a modal expressing a higher probability and the one expressing a lower probability, they are oblivious to the semantic differences caused by the different ordering of the two modals in a nested expression.

General discussion

In the above three experiments, we examined the processing of nested epistemic modality focusing on how interlocutors interpret modally nonharmonic expressions, such as "He *certainly may* have forgotten." Given that the expression contains two modals with contrasting epistemic strengths, it is natural to ask how interlocutors interpret the meaning of it. According to current linguistic theories, the meaning of the second modal should be interpreted within the scope of the first modal (Lyons, 1977; Moss, 2015; Potsdam, 1998). Based on this account, if the first and second modals switch their positions, a change in meaning should be expected. However, it is also possible that in casual conversation, the scope of nested modals may not be thoroughly processed, and thus the order of the modals may not significantly change interlocutors' interpretation of the meaning of the nested expression. Since research on the processing of nested epistemic expressions is so limited, it remains an open question whether in everyday situations interlocutors process the meaning of nested epistemic expressions according to the linguistic representations that are assumed to underlie the forms.

The investigation of meaning can be approached from different angles with different levels of depth (Putnam, 1975), and in this study we focused specifically on the epistemic strength of the expression, the degree of probability indicated by epistemic modals. Following Degen et al. (2019), Renooij and Witteman (1999), and Willems et al. (2019), we treated the strength of the epistemic modals as a property quantifiable on a continuous scale, and thus individuals' knowledge about the strength of the modals could be elicited by using the probability rating task. We found that the statement embedded within an epistemic expression of a higher epistemic strength were rated higher on the scale of probability than the same statement embedded within an epistemic expression of a lower strength, which indicated the sensitivity of the rating task to the strength of the epistemic expression. The immediately following question is when an expression is embedded within two epistemic modals, how does the probability rating of that expression reflect the meaning of the nested epistemic modals?

Though Lyons (1977), Moss (2015), and Potsdam (1998) did not further specify how the scopes of the component modals contribute to the overall epistemic strength of the expression, we argued that the strength of the nested epistemic expression is a function that applies to the strength of each component modal, and the computational process is further influenced by the anchoring effect. To be specific, if the scopes of the component modals have been thoroughly processed, an arithmetic computation will take place with the strength of each component modal as the input and the overall

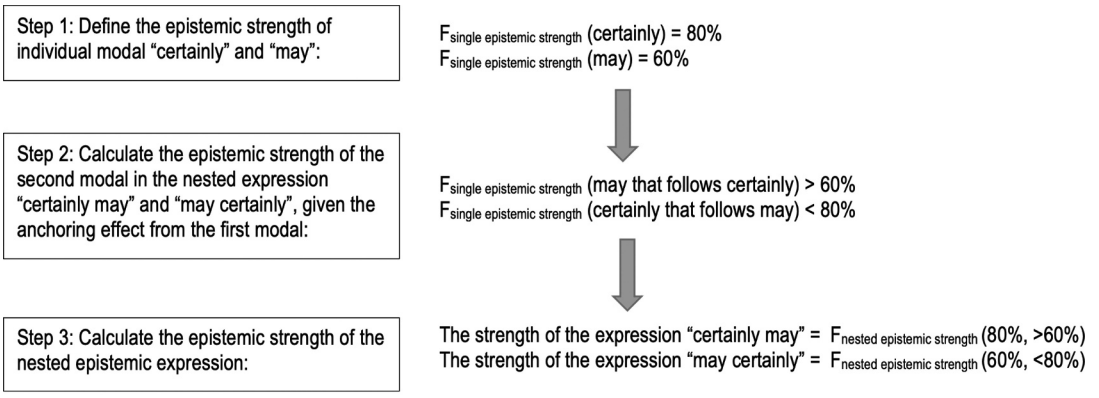

Figure 11. The computation of epistemic strength based on the scope account. The capital letter "F" stands for a function, with its argument(s) included in the parenthesis.

strength of the expression as the output. As illustrated in Figure 11, suppose the strength of individual modal *certainly* and *may* is 80 and 60 out of 100, respectively. When those two modals co-occur in the same clause and the scopes of both modals have been carefully processed, the anchoring effect of the first modal changes the epistemic strength of the second modal. In the case of *High-Low* word order such as *certainly may*, the strength of the second modal *may* is higher than 60% due to the anchoring effect of the modal *certainly*. On the other hand, in the case of *Low-High* word order such as *may certainly*, the strength of *certainly* is lower than 80% due to the anchoring effect of *may*. Thus, when calculating the overall strength of the nested expressions, the expressions with different word orders feed different sets of input to the function. Given that the sum of the individual modal strength in the *High-Low* condition is larger than that of the *Low-High* condition, we predicted that *certainly may* would be rated higher than *may certainly*. This prediction is based on our intuition that the sum of the strength of the component modals and the overall strength of the nested expression are correlated to some extent. However, the function that calculates nested epistemic strength ($F_{\text{nested epistemic strength}}$) under the scope account we illustrated above is unlikely to be a simple operation like multiplication or addition; otherwise, the rating of the nested-modal conditions will be lower than both of the single-modal conditions or higher than both of them, which was not what we found in this study. By comparison, the good-enough processing account predicts that the processor is not sensitive to the order of the nested modals, following a simpler algorithm: If the strength of the second modal is not the same as that of the first modal, the output will be the average of the individual epistemic strength, which suggests that the processor treats the occurrence of two contrasting modals as an indication of uncertainty rather than fully pin down the scope of the two modals or compute the meaning of the second modal under the influence of the first modal. The findings of the three experiments we conducted revealed patterns more consistent with the good-enough processing account.

In the first experiment, participants judged the probability of a statement *S*, given either "modal1 modal2 S" or "modal2 modal1 S". The two epistemic modals, *modal1* and *modal2*, were selected from different positions on the scale of epistemic strength (Halliday & Matthiessen, 2004; Holmes, 1982) so that the nested epistemic expressions studied in this experiment were truly nonharmonic (Lyons, 1977). What we found was that the order of the two modals did not affect the probability rating score of the statement. The difference between the two nested modal conditions in probability rating was about 2 out of 100, which was too small to be statistically meaningful. Moreover, significant differences in probability rating scores were observed between statements in the nested conditions and single modal conditions, suggesting that the epistemic strength of the nested expressions was in between the

epistemic strength of the two component modals. Same patterns were observed in the second experiment, in which *modal1* and *modal2* were selected from the high and low extremes on the scale of epistemic strength. If there was a chance that in the first experiment the difference in epistemic strength between the two modals was too small to trigger an order effect, it shouldn't be the case for the second experiment, in which one modal expressed very high probability (*certainly* and *definitely*), while the other expressed very low probability (*may* and *might*). Results from the first two experiments suggested that when receiving a statement embedded within two epistemic modals, one expressing high probability and the other expressing low probability, the interlocutor would assign to the embedded statement a medium probability in between the epistemic strength of the individual modals, regardless of the order of the two modals.

In the third experiment, parenthetical elements were inserted in between the two modals in the nested expressions, so that the two modals were no longer adjacent to each other like they were in the previous two experiments. If the reason for not finding a word order effect in previous experiments was due to participants interpreting nested modals together as one idiomatic item (like the case of *might could* in some dialects of American English), this should not be the case for the third experiment. Still we found that the order of the two modals in the nested expressions did not affect the probability rating of the statement, even when the two modals were not adjacent to each other. The explanation we offered for the lack of order effect was that when readers encountered two epistemic modals in a sentence, no matter whether the modals were adjacent or separated, readers would treat the occurrence of the two modals as an indicator of uncertainty, without further pinning down the scope of each modal. This explanation echoes the good-enough processing theory of language processing (Christianson et al., 2001, 2006; Ferreira et al., 2001).

Moreover, the third experiment revealed an interesting pattern that seems to suggest a possible interaction between the inclusion of parenthetical elements and the epistemic strength of single epistemic modals. For the previous two experiments in which the parenthetical elements were not included, significant differences were found between the single high-modal condition, single low-modal condition, and the nested-modal conditions. To be more specific, ratings of the statements in nested conditions were in between the ratings of the two single-modal conditions. However, for the third experiment in which parenthetical elements were included, the rating scores of the nested-modal conditions and the single low-modal condition were essentially the same.

The parenthetical elements in the third experiment were evidential expressions indicating the source of information based on which the statements were made. Rooryck (2001) discussed how parentheticals in English, such as *I think* and *they say*, convey a variety of evidential meanings. An interesting observation was that in an evidential parenthetical, the meaning of the verb is generally impoverished. For example, in sentence "This building, I'm afraid, is going to be demolished", the parenthetical *I am afraid* expresses not so much the fright of the speaker as an emotional status but "a reluctant statement of probable fact" (Rooryck, 2001, p. 128). Similarly, in the utterance "Jules will be late, he said," the parenthetical, *he said*, does not mean that the statement *Jules will be late* is exactly what *he* said. It is the speaker of the utterance that makes an assertion that Jules will be late, and to further support this assertion, the speaker adds the source, based on which the assertion is made. It is possible that what *he said* was not a direct assertion of Jules' lateness but information consistent with such assertion, such as that Jules' car got a flat tire. The function of the parenthetical *he said* is similar to *I think* or *probably*, which is an epistemic expression of probability (Reinhart, 1983, p. 175).

If parentheticals are also epistemic expressions of probability as Rooryck (2001) suggested, it follows that the inclusion of parenthetical elements not only enlarged the distance of the modals, but also influenced participants' rating of individual experimental items. However, since the parenthetical element remained the same across all conditions of the same item and varied across different items, if comprehenders processed the two nested conditions differently, such difference should be captured by this research paradigm when we averaged the rating scores across all the items of the same condition. Instead of finding an order effect among the two

nested-modal conditions, what we found in this experiment was that items with only a single modal of low epistemic strength were rated much higher than before, and there was no statistical difference between the nested-modal conditions and the single low-modal condition. This result suggested an interesting interaction between the parentheticals and the modals of low epistemic strength. It seems that the presence of the parenthetical elements increased the credibility of the statement embedded within a single modal of low epistemic strength. For example, comparing the utterance "Based on my experience, the soup may have some star anise in it" with "The soup may have some star anise in it," the result of this experiment suggests that comprehenders would assign higher probability to "the soup has star anise" when they hear the first sentence rather than the second sentence.

Linguistic theories have cataloged the morphological system of evidentiality across languages and established the hierarchy of information source based on its credibility. For example, direct witness is regarded as the most reliable source of information across languages while statements based on assumptions are perceived as having the lowest credibility (Song, 2018). Unfortunately, few research attempts have been made to illustrate how in general the inclusion of parenthetical evidentials influences the credibility of a statement within the scope of epistemic modals. Drawing on the insight of Reinhart (1983), we postulate that, in general, an evidential parenthetical in English expresses median degree of probability similar to the epistemic strength of the modal *probably*. Admittedly, comprehenders will assign a very high probability to the statement "the soup has star anise" when they hear "according to the cook, the soup may have some star anise in it," while assigning a very low probability to the same statement when they hear "according to a child who always lies, the soup may have some star anise in it." However, if the knowledge background of the source is obscured, the default credibility of an evidential parenthetical in English is medium degree of probability. For example, in the utterance "Based on my experience, the soup may have some star anise in it," it is not clear whether the speaker is as experienced as the cook, or as dishonest as the child. In this case, the comprehender tends to treat the parenthetical as expressing a median degree of credibility. Given that the epistemic modal *may* in this sentence expresses low probability, the inclusion of the parenthetical boosts the overall probability of the statement.

Interestingly, the inclusion of the parenthetical elements did not seem to affect the single high modal conditions to the same extent as the single low modal condition. As we can see from Tables 7 and 8, the rating score of the single high modal condition was much higher than that of the other conditions. However, to statistically examine whether or not there is an interaction effect between the inclusion of parenthetical elements and epistemic modals, we need a different research design that includes both modal conditions and parentheticals as fixed effects. Since the research question of this project is on the processing of nested epistemic expressions, the preliminary finding pertaining the processing of parenthetical elements will be addressed in future projects.

The focus of this study is to investigate whether or not the processing of nested epistemic expressions follows the prescription of formal linguistic theories, which suggested that the second modal in a nested expression should be interpreted within the scope of the first modal (Lyons, 1977; Moss, 2015; Potsdam, 1998). The notion of scope is one of the fundamental and most frequently discussed concepts in the study of language and logic, which defines the sequence of logical operations necessary for the generation of meaning (Hintikka, 1997). The scope account of nested epistemic expression interprets sentence "He certainly may have forgotten" as the equivalent of *it is certainly the case that he may have forgotten*. By contrast, "He may certainly have forgotten" is interpreted as the equivalent of *it may be the case that he certainly has forgotten*. The two expressions have different meanings because the modal operators differ in their priority. In the first expression *may* is embedded within the scope of *certainly*, while in the second expression *certainly* is embedded within the scope of *may*. If in daily conversation, interlocutors process the meaning of nested epistemic

expressions following the exact logical operation sequence as stipulated by the scope account without any interference from other cognitive domains, we would find that interlocutors interpret the meaning of nested epistemic expressions differently depending on the order of the two modals.

What we actually found in this study was the absence of the order effect regardless of whether or not the two modals were adjacent or separated by parenthetical elements. This finding echoes with the well-documented semantic illusions (Barton & Sanford, 1993; Erickson & Mattson, 1981; Otero & Kintsch, 1992) as well as the good-enough processing framework (Christianson et al., 2006; Ferreira et al., 2002, 2001; Ferreira & Lowder, 2016) which highlights the fact that the parser performs superficial analysis of linguistic input based on heuristics, leading to inaccurate interpretations. Although findings in both semantic illusion literature and this study reveal how processing diverges from the linguistic representation of the input, we believe that there are different mechanisms that underlie the two processes. Semantic illusions such as the case of "bury the survivors" and the "Moses illusion" are largely due to the strong prior belief of the context (Otero & Kintsch, 1992). While for this study, the lack of order effect during the processing of nested epistemic expressions results from the mixture of contradictory epistemic strength of the two modals. The epistemic strength of the first modal lingers in memory, and is mixed with the epistemic strength of the second modal when it is encountered. The parser then normalizes the mixed epistemic strength as an indicator of general uncertainty without further pinning down the scope of each modal.

Conclusion

The findings of this study suggest that when processing "modally nonharmonic" nested epistemic expressions, such as "he *certainly may* have forgotten," interlocutors generally interpret the occurrence of two modals in one clause as an expression of uncertainty, which is different from the epistemic strength of each component modal. Moreover, changes in word order do not seem to change interlocutors' interpretation of the epistemic strength of the expression, and this pattern holds true whether or not the two modals are separated in the string. This finding challenges the scope account of nested epistemic expressions in language processing, and suggests a holistic processing mechanism in line with the "good-enough" processing framework.

Notes

1. https://github.com/PON2020/Nested_Modality_Twitter.
2. Based on Reinhart (1976), node A c(onstituent)-commands node B iff the branching node most immediately dominating A also dominates B.
3. https://github.com/PON2020/Nested_Epistemic_Expressions_Submission/tree/main/Stimuli.
4. 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.
5. https://github.com/PON2020/Nested_Epistemic_Expressions_Submission/tree/main/Data_Analysis_Nested_Expression.
6. Fifty weakly informative priors were selected from the normal distribution with a variety of mu and sigma combinations. Ninety percent of the result we obtained from the Bayes factor analyses using those priors were in favor of the null hypothesis. The script and output of the reported Bayes factor analyses are available in GitHub (https://github.com/PON2020/Nested_Epistemic_Expressions_Submission/tree/main/Data_Analysis_Nested_Expression).
7. Fifty weakly informative priors were selected from the normal distribution with a variety of mu and sigma combinations. Ninety-six percent of the result we obtained from the Bayes factor analyses using those priors were in favor of the null hypothesis. The script and output of the reported Bayes factor analyses are available in GitHub following the same link in the note above.
8. Fifty weakly informative priors were selected from the normal distribution with a variety of mu and sigma combinations. All the Bayes factor analyses using those priors showed results in favor of the null hypothesis.

Acknowledgments

We thank Dr. Emily Morgan for her suggestions on mixed-effects models used in this study and Dr. Kenji Sagae for suggesting the corpus study reported in this article. We also thank Dr. Adam Sennet for his insightful comments on this project and all the undergraduate research assistants who helped with this project.

Disclosure statement

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

Funding

This work was supported by 1R01 HD100516-01A1; NIH/NICHD.

ORCID

Zhuang Qiu http://orcid.org/0000-0002-1726-6214
Fernanda Ferreira http://orcid.org/0000-0002-9349-9300

References

Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford University Press.
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. https://doi.org/10.3758/BF03197179
Battistella, E. L. (1995). The syntax of the double modal construction. *LINGUISTICA atlantica*, *17*, 19–44. https://journals.lib.unb.ca/index.php/la/article/view/22489
Bürkner, P.-C. (2017). brms: An r package for Bayesian multilevel models using stan. *Journal of Statistical Software*, *80* (1), 1–28. https://doi.org/10.18637/jss.v080.i01
Bybee, J. L. (1985). *Morphology: A study of the relation between meaning and form*. (Vol. 9). J. Benjamins.
Chafe, W. L., & Nichols, J. (Eds.). (1986). *Evidentiality: The linguistic coding of epistemology*. (Vol. 20). Ablex Pub. Corp.
Christianson, K., Hollingworth, A., Halliwell, J. F., & Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognitive Psychology*, *42*(4), 368–407. https://doi.org/10.1006/cogp.2001.0752
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. https://doi.org/10.1207/s15326950dp4202_6
Coates, J. (1983). *The semantics of the modal auxiliaries*. Routledge.
de Haan, F. (2001, april). The place of inference within the evidential system. *International Journal of American Linguistics*, *67*(2), 193–219. Retrieved 2021, August 13, from https://doi.org/10.1086/466455
De Swart, H. (1998). Licensing of negative polarity items under inverse scope. *Lingua*, *105*(3–4), 175–200. https://doi.org/10.1016/S0024-3841(98)00021-7
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. https://doi.org/10.1016/j.pragma.2018.11.015
Di Paolo, M. (1989). Double modals as single lexical items. *American Speech*, *64*(3), 195–224. https://doi.org/10.2307/455589
Elsman, M., & Dubinsky, S. (2009). Double modal syntactic patterns as single modal interactions. *University of Pennsylvania Working Papers in Linguistics*, *15*(1), 10. https://repository.upenn.edu/cgi/viewcontent.cgi?article=1076&context=pwpl
Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological Science*, *17*(4), 311–318. https://doi.org/10.1111/j.1467-9280.2006.01704.x
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. https://doi.org/10.1016/S0022-5371(81)90165-1
Ferreira, F., Christianson, K., & Hollingworth, A. (2001). Misinterpretations of gardenpath sentences: Implications for models of sentence processing and reanalysis. *Journal of Psycholinguistic Research*, *30*(1), 3–20. https://doi.org/10.1023/A:1005290706460
Ferreira, F., Bailey, K. G., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science*, *11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Ferreira, F., & Lowder, M. W. (2016). *Psychology of learning and motivation* ,*65*, 217–247. https://doi.org/10.1016/bs.plm.2016.04.002
Gisborne, N., & Holmes, J. (2007, march). A history of English evidential verbs of appearance. *English Language and Linguistics*, *11*(1), 1–29. https://doi.org/10.1017/S1360674306002097
Halliday, M. A. K., & Matthiessen, C. M. I. M. (2004). *An introduction to functional grammar* (3rd ed). Arnold; Distributed in the United States of America by Oxford University Press.
Hintikka, J. (1997). No scope for scope? *Linguistics and Philosophy*, *20*(5), 515–544. https://doi.org/10.1023/A:1005346511016
Holmes, J. (1982). Expressing doubt and certainty in English. *RELC Journal*, *13*(2), 9–28. https://doi.org/10.1177/003368828201300202
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. https://doi.org/10.1016/S1060-3743(97)90033-3
Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, *21*(11), 1161–1166. https://doi.org/10.1177/01461672952111004
Kratzer, A. (2012). *Modals and Conditionals: New and Revised Perspectives*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199234684.001.0001
Ladusaw, W. A. (1979). *Polarity sensitivity as inherent scope relations*. The University of Texas at Austin.
Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.
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. https://doi.org/10.3765/sp.8.5
Nagle, S. J. (1994). The English double modal conspiracy. *Diachronica*, *11*(2), 199–212. https://doi.org/10.1075/dia.11.2.04nag
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. https://doi.org/10.1515/9783110895339.349.
Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas—part ii. *Language and Linguistics Compass*, *10*(11), 591–613. https://doi.org/10.1111/lnc3.12207
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. https://doi.org/10.1111/j.1467-9280.1992.tb00034.x
Perkins, M. R. (1983). *Modal expressions in English/michael r. perkins*. F. Pinter London. Book. 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).
Putnam, H. (1975). The meaning of meaning. *Minnesota studies in the philosophy of science*, *7*, 131–193. https://conservancy.umn.edu/handle/11299/185225
Reinhart, T. M. (1976). *The syntactic domain of anaphora* [Unpublished doctoral dissertation]. Massachusetts Institute of Technology.
Reinhart, T. M. (1983). Point of view in language: The use of parentheticals. In G. Rauh(Ed.), *Essays on Deixis* (pp. 169–194). Tübingen: Gunter Narr Verlag.
Renooij, S., & Witteman, C. (1999). Talking probabilities: Communicating probabilistic information with words and numbers. *International Journal of Approximate Reasoning*, *22*(3), 169–194. https://doi.org/10.1016/S0888-613X(99)00027-4
Rooryck, J. (2001). State of the article: Evidentiality part i. *GLOT International*, *5*(4), 125–133. https://hdl.handle.net/1887/14555
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. https://doi.org/10.1016/j.jml.2013.04.001
Song, J. J. (2018). *Linguistic typology*. Oxford University Press.
Traxler, M. J. (2011). *Introduction to psycholinguistics: Understanding language science*. Wiley-Blackwell.
Traxler, M. J. (2014). Trends in syntactic parsing: Anticipation, Bayesian estimation, and good-enough parsing. *Trends in Cognitive Sciences*, *18*(11), 605–611. https://doi.org/10.1016/j.tics.2014.08.001
Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124
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. https://doi.org/10.3102/1076998610396895
Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the savage–dickey method. *Cognitive Psychology*, *60*(3), 158–189. https://doi.org/10.1016/j.cogpsych.2009.12.001
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*. https://doi.org/10.48550/arXiv.1901.09686