Processing Nested Epistemic Expressions

By

ZHUANG QIU

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Linguistics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Fernanda Ferreira , Chair

Emily Morgan

Kenji Sagae

Committee in Charge

2022

Abstract

In this dissertation, I explored the semantic processing of a construction that has been largely uninvestigated, namely, the nested structure of two epistemic modals in a single clause, as illustrated in the sentence "He may certainly have forgotten" (Lyon, 1977). There are two different theoretical approaches that may account for the processing of this structure. The formal semantics account (Lyons, 1977; Potsdam, 1998; Moss, 2015) claims that the meaning of the inner modal should be interpreted within the scope of the outer modal. Based on this account, if the first and second modals switch their positions, a change in meaning should be expected. This account is also referred to as the "scope account". In contrast, a good-enough processing account (Ferreira & Lowder, 2016) predicts that the scope of nested modals may not be thoroughly processed, and thus, the order of the modals may not change interlocutors' interpretations of the nested expression. Using a combination of Bayesian modeling and judgment tasks, in six experiments I evaluated these two theoretical perspectives, and the result suggested a holistic processing mechanism in line with the good-enough processing framework. This dissertation consists of five chapters. The first chapter introduces the research questions and outlines the structure of this dissertation. The second chapter provides an overview of the research background where my dissertation project is situated. The third chapter reports six experiments I conducted examining how interlocutors process nested epistemic expressions in casual conversations, focusing on the extent to which the prediction of the scope account matches the patterns observed in the experiments. The fourth chapter discusses the major findings that have been consistently replicated in the six experiments, and proposes a possible account for the cognitive mechanism underlying the processing of nested epistemic expressions. The last chapter concluded this dissertation, summarizing the answers to the research questions.

To my parents, Kong Wen, Qiu Lujun

To my grandma, Kong Xiangzhen

To my hamster, Li Shuyao on her twenty-third birthday

Acknowledgement

I want to thank my mom and dad, to whom I owe a great debt of gratitude for their lifelong hard work that supports the family, making every achievement in my life possible!

Special thanks to my superb mentor Fernanda Ferreira for all the guidance, care and support ever since I joined the lab! Thank you for encouraging me and standing up for me during those challenging times, and thank you for making me a psycholinguist!

Thanks to Emily Morgan and Kenji Sagae for your valuable feedback that improved the manuscript considerably; and to Adam Sennet, John A. Hawkins, David Corina, Raul Aranovich and John M. Henderson for your comments and suggestions during the early stage of the project.

Thanks to my colleagues and undergraduate research assistants in the linguistics department and the Ferreira Lab. I am grateful for all your effort in creating an inclusive, intellectual, and caring environment for learning and doing research. Thanks to the staff members in the linguistics and psychology program, especially Stephanie Fallas.

Thanks to Thieng Nguyen, my primary care doctor at the school hospital.

Thanks to Suphasiree Chantavarin a.k.a Nene for being not only a friend but also a role model of mine! It is encouraging to be close to someone that carries many virtues that I admire: wisdom, compassion, conscientiousness...Thanks to Bruno Debus for being such a wonderful neighbor and friend. I feel privileged to be invited into your little bubble of simplicity and tranquility. Thanks Selina, Serena, Leo, Tim, and Rita for saving me from work burnout!

My special thanks to Dave and the Corina family for your kindness and care during the holiday season of 2020. The Christmas tree will light up whenever I feel lonely and cold.

At last, to Big Haru Lea, thanks for being my hamster!

TABLE OF CONTENT

1 Introduction

Suppose you plan to go hiking tomorrow and hear your friend say "it *might* be raining tomorrow". Will you bother cancelling 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 human beings 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". 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 "It *may certainly* rain tomorrow" has two modals, which are "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 conflict in their epistemic strength. 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 non-harmonic” combination as opposed to “modally harmonic expression”, like “he may possibly have forgotten” in which, both the word “may”, and “possibly” express low probability (Lyons, 1977, p. 807).

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 frequent people use nested epistemic expressions in daily communication, we created a database of 413986 tweets, each containing at least one epistemic modal, and searched for nested epistemic expressions in the database. In total, we found about 4200 tweets containing nested expressions, with half of them being harmonic expressions like “may possibly”, and the other half being non-harmonic 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 has been argued that in a non-harmonic nested epistemic expression, one

[1] https://github.com/PON2020/Nested Modality Twitter.git

modal must be within the scope of the other (Lyons, 1977, p. 808; Moss, 2015; Potsdam, 1998). Based on this account, the modally non-harmonic 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", though the two utterances are identical except for the order of the two 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 sentence processing (Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Lowder, 2016), on the other hand, suggested that the parser may perform superficial analysis of linguistic input based on heuristics, leading to inaccurate interpretations. Based on this account, 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". 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.

In this dissertation, I reported six experiments studying the cognitive mechanism of processing nested epistemic expressions, especially focusing on the extent to which the scope of the modals is processed by interlocutors during informal conversations. First of all, I reviewed the research background of this research project, which formed the second chapter of my dissertation. The chapter started with an introduction to different notional categories of modality, focusing on the semantics of epistemic modality and its relation to other notional categories like deontic modality. An important shared feature between epistemic and deontic modality is that they can be viewed as a quantitative scale, carrying both semantic and pragmatic implications that have been attested by the findings of some experiments in this project. The chapter then narrowed down to the research topic, a specific way of expressing epistemic modality called

nested epistemic expressions. I clarified the definition of nested epistemic expression, distinguishing it from another similar linguistic construction used in southern American dialects, and reported a preliminary corpus research on the use of nested epistemic expressions in online social network. After that, I contrasted two theoretical accounts for the processing of nested epistemic expressions, namely, the scope account and the good-enough processing account, discussing their underlying assumptions and their different predictions in terms of how people interpret the strength of nested epistemic expressions.

The third chapter reported six experiments examining the extent to which the predictions of the scope account and good-enough processing account match the way people interpret nested epistemic expressions in experimental settings. Experiment one to three used the same single-factor design with four experimental conditions to explore whether or not the order of the component modals affects how people interpret the strength of the nested expression. The relative strength of the component modals and the distance between them varied across the three experiments, however, none of these experiments found the order of the nested modals influence how participants rated the probability of the nested expressions. Experiment four and five adopted a two factor design, examining possible interaction between the order of the component modals and the distance between them on how people process nested epistemic expressions. The paradigm of the two experiments were similar in most aspects, except that the stimuli in Experiment four were presented visually while the same set of stimuli were presented acoustically in Experiment five. Though an interesting interaction between modal strength and parenthetical elements were observed in single modal conditions, the order of the modals and the distance between them did not affect the way people processed nested epistemic expressions. The last experiment reported in Chapter three adopted a new paradigm in which participants

were provided with two interpretations of a nested expression, and were asked to choose the option that is closer to their own interpretation. Though using a different research paradigm, this experiment successfully replicated the findings of the previous experiments.

Chapter four discussed the patterns in processing epistemic expressions that were consistently observed in all six experiments, and proposed a possible account for the cognitive mechanism underlying the processing of nested epistemic expressions. Some of the research findings brought insights to the nature of evidentiality and logical entailment, and were also analyzed in detail. This dissertation ended with a short conclusion, answering the research questions and pointing out a possible direction for the future research.

References

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.

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

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

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

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

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

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

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

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.

Kratzer, A. (2012). The Notional Category of Modality. *In Modals and Conditionals*. Oxford: Oxford 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, 5–1.

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

Perkins, M. R. (1983). *Modal expressions in English*. London: Pinter. Retrieved from http://www.loc.gov/catdir/toc/fy0605/88202602.html

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

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

2 Research background

2.1 The ontology of modality

Modality is a cross-language grammatical category, which is “concerned with the status of the proposition that describes the event” (Palmer, 2001 :1). Similar to tense and aspect, modality is a verbal property, operating on the clause level; however, modality is different from tense and aspect in that it does not refer to the temporal characteristics of the event. Rather, modality is related to speakers' perception or attitude towards the event.

Functional theories categorize modality into inter-related subdivisions based on their respective semantic and pragmatic functions. According to Palmer (2001), mood and modal systems are two major grammaticalized categories of modality, each focusing on different but not always distinct aspects of modality. Mood is “an inflectional representation of modality” (Collentine, 2010), prototypically binary, contrasting realis events (indicative mood) with irrealis events (subjunctive mood). The modal system, on the other hand, is concerned with speakers’ commitment to the truth-value of a proposition or speakers' will to carry out a potential event. The former is referred to as propositional modality while the latter is called event modality. Each of the above modal systems can be further divided into subcategories. Propositional modality has two major subcategories, evidential modality indicating different information sources from personal experience to hearsay (Aikhenvald, 2014) and epistemic modality indicating the likelihood of a proposition being true (Kratzer, 1981). As to event modality, it can be further divided into deontic modality and dynamic modality. The former pertains to will, obligation, and permission (Traugott, 1989: 32), while the later relates to ability and willingness (Palmer, 2001 :10).

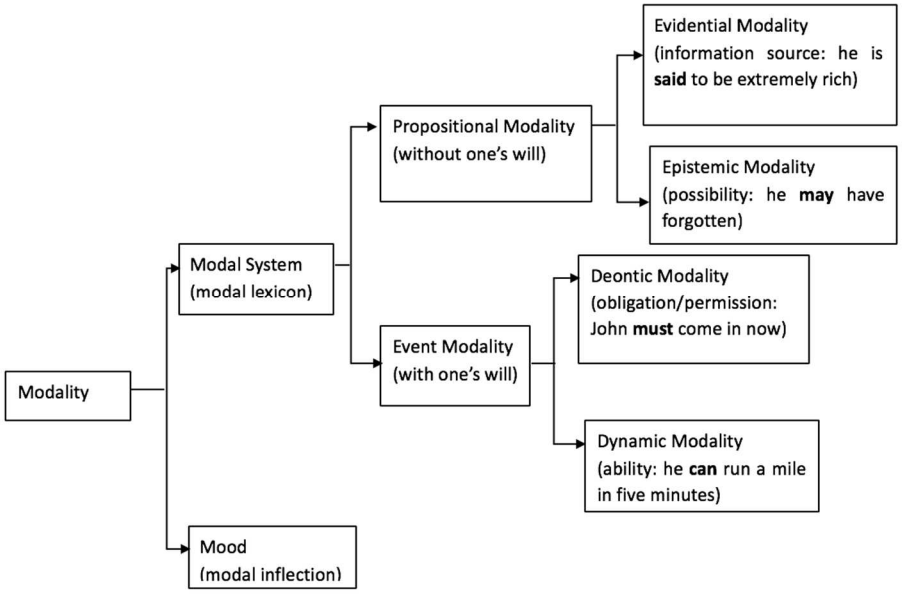


Figure 2-1 A summary of the categorization of modality based on Palmer (2001)

A brief summary of the categorization of modality is presented in Figure 2-1, and it is important to note that modality is far more complex than the seemingly neat branching represented above. In fact, cross-linguistically, the boundaries between different types of modality are not always clear, and there are also language specific modal systems (Palmer, 2001:10-14). However, in terms of epistemic modality, its definition and the scope of analysis are less disputed. Though the focus of this research project is the semantics of epistemic modality and not the other modal categories, the functional ontology of modality provides an overview showing where the research topic is situated in the system of modality.

From the perspective of generative grammar (the minimalist program in particular), mood/modality is analyzed as the head of some functional categories. Ager (2003) suggested that a modal item is the head of TP, which takes a clause as its complement. Different from other

lexical heads (such as verbs, nouns, and prepositions), the head of TP does not assign any theta-roles, and thus is called a functional category. According to Ager (2003), the marking of tense, aspect and modality in English is the consequence of merging and feature checking within the T head. Cinque (2006), on the other hand, proposed multiple functional categories to accommodate the inflectional features related to tense, aspect and modality. These functional categories follow a rigid hierarchy, and each of the categories checks off one inflectional feature as it merges with the VP. The highest functional categories in the hierarchy are related to mood or modality, with the top ten being "MoodP (speech act) > MoodP (evaluative) > MoodP (evidential) > ModP (epistemic) > TP (Past) > TP (Future) > MoodP (irrealis) > ModP (alethic) > AspP (habitual) > AspP (repetitive)" (Cinque, 2006:12). These categories are supposed to account for linguistic properties related to tense, aspect and modality across all languages, though one specific language can only have a proportion of those categories. As can be seen from the above review, functional linguists analyze the meaning of various modal expressions focusing on its contribution to the discourse, while generative linguists pay more attention to the syntactic property of the modals, trying to provide theoretically coherent account for syntactic operations between modals and other lexical or functional categories.

2.2 Epistemic modality

Epistemic modality indicates speakers' commitment to the truth value of what is said (Kratzer,1981; Coates, 1983), and it serves as an important means to modify the strength of the argument (Hyland & Milton, 1997). It can be expressed in various ways, one of which is through the use of lexical items, such as modal auxiliaries, adjectives and adverbs. This paper uses the term "epistemic modals" to refer to the words that express epistemic modality. Kratzer (2012)

argues that the semantics of modals, including epistemic modals, has two important aspects. One is called the modal base, which is a set of propositions that represent the body of evidence, and the other is called the ordering source, which is a set of propositions that represent the ideal in the world of evaluation. According to Kratzer, a proposition is possible if and only if its negation is not a necessity in any world with respect to the modal base and ordering source (Kratzer, 2012: 40). In other words, the available evidence defines a set of possible worlds, which are ranked according to how close they are to the norm in the world of evaluation. A proposition is possible if and only if among the worlds that are consistent with the body of evidence, there is at least one world in which that proposition is true.

The framework of modal base and ordering source inspires a large number of studies on different types of modality such as the study of generics (Cohen, 2012), and deontic modality (Björnsson & Shanklin, 2014). However, the concepts of modal base and ordering source are hard to operationalize in an empirical research design. Moss (2015), on the other hand, offers an alternative perspective to conceptualize the meaning of epistemic modality, which is comparatively easier to test with empirical evidence. According to Moss (2015:29), the semantics of epistemic expressions can be modeled as having mental committee members vote on the acceptance of a statement. For example, "must S" means "every committee member accepts S", while "might S" means that "some committee member accepts S". This framework was adopted in the design of Experiment six reported in Chapter three.

It worth noting that the term "epistemic modals" used by logicians denotes a variety of syntactic categories, such as adverbs, adjectives, and auxiliary verbs. Among those lexical categories, this study pays special attention to epistemic adverbs and auxiliaries. Potsdam's (1998) syntactic analysis of adverbs sheds light on the structural aspect of the epistemic modals

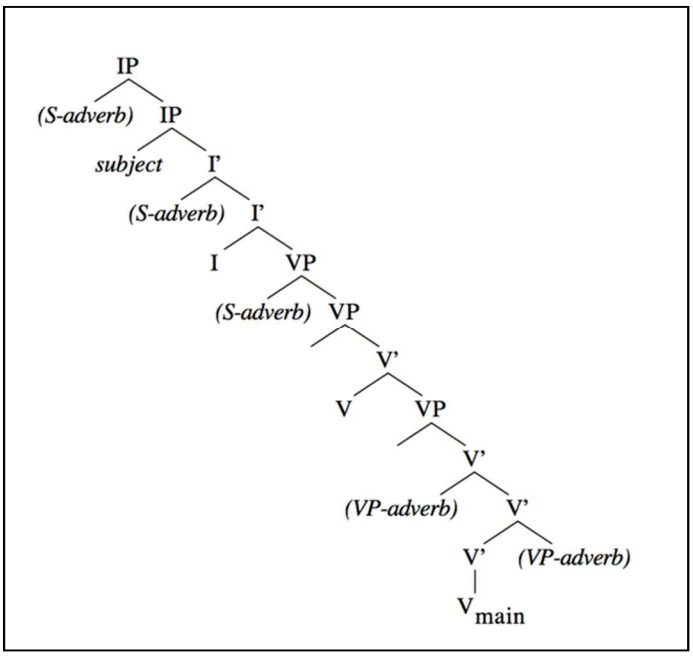


Figure 2-2 Potsdam's proposal on the syntax of S-adverbs and VP adverbs

investigated in this study. Potsdam argued that the difference between S-adverbs (sentence modifiers such as epistemic adverbs) and VP-adverbs (verb modifiers such as "carefully", "quickly") can be syntactically explained by assigning them to different positions in the tree structure shown in Figure 2-2. According to his proposal, a VP-adverb is syntactically lower in the tree than does an S-adverb, sitting immediately adjacent to the main verb. On the other hand, S-adverbs are high up in the syntactic structure, being left adjoined to IP, I' or the topmost VP. Since epistemic adverbs are S-adverbs, and the epistemic auxiliaries belong to the category of I (Chomsky, 1957; Emonds, 1976), Potsdam's proposal provides the syntactic framework for this research. The nested epistemic modals examined in this research are cases in which the epistemic adverbs are left adjoined to I' or the left adjoined to the topmost VP, like the word "certainly" in sentence "he certainly may have forgotten" and "he may certainly have forgotten" respectively. Moreover, the scope difference between different epistemic items can also be reflected in the

different syntactic positions they occupy, which will be further discussed in the section of nested epistemic vocabulary.

Linguists also propose theories focusing on the degrees of probability conveyed by different epistemic lexical items. This study uses the term "epistemic strength" to refer to the degree of probability expressed by epistemic modals. Horn (1972) analyzed the epistemic strength of various modals, placing them in different positions on scales such as "certain-probable/likely-possible", and "uncertain-chancy-improbable-impossible". Halliday (1970) argued that the basic distinction in the degree of probability is the distinction between "probable" and the polar values, "possible" and "certain". This idea is echoed by Holmes (1982) who proposed a three-point scale from "certain" to "probable" to "possible". Halliday and Matthiessen (2004) further specified the lexical items that fall into each of the three categories. For example, "certain (certainly)" and "must" express high degree of probability, "probable (probably)" and "will (would)" express median probability, while "possible (possibly)" and "may (might)" express low probability (116; 622). The three-point scale of epistemic strength has been adopted in a considerable amount of empirical research on epistemic expressions, such as the research on the semantics and pragmatics of modals (Degen et al., 2019; Willems, Albers, & Smeets; 2019), and the research on the use of epistemic modals in academic writing (McEnery & Kifle, 2002; Hu & Li, 2015).

Items on the same scale are not only qualitatively similar (Gazdar, 1979) and quantitatively comparable; they also carry specific semantic and pragmatic implications, as Levinson (1983: p.134) nicely summarized "the semantic content of lower items on a scale is compatible with the truth of higher items obtaining, and the inference that higher items do not in fact obtain is merely an implicature". Semantically speaking, an utterance containing an element

higher on the scale entails an utterance containing an element lower on the scale (Horn, 1972; Gazdar, 1979; Van der Auwera, 1996). For example, the epistemic modal "certain" is higher on the epistemic scale than "possible", and the sentence "it is certain that he is late" logically entails that "it is possible that he is late", but not the other way round.

The entailment relation shows that the two utterances "it is certain that he is late" and "it is possible that he is late" can be both true at the same time, however, pragmatically speaking, the utterance containing the lower item on the scale has an implicature that the utterance with a higher item does not hold. In the above example, saying "it is possible that he is late" implies that the speaker does not think it is certain that he is late. Such implicature arises from the Gricean maxim of quantity which stipulates that conversational interlocutors should make their exchange as informative as required for the purpose of the conversation (Grice, 1975; 1978). Items higher on a scale is more informative than items lower on the scale (Gazdar, 1979; Verstraete, 2005), meaning that the sentence "it is certain that he is late" contains more information than "it is possible that he is late". Thus, in the context when both of these two sentences are true, the first sentence is always preferred by the speaker. It follows that to select the second sentence rather than the first sentence implies that the speaker thinks the first sentence is not true. The above mechanism of conversational implicature explains why in daily conversation we process the meaning of an epistemic modal based on the face value of it, rather than the logically entailed values of it. So when we were told in a lottery that we "may win a 5000 dollar cash prize", we knew that it was not the case that we "are certainly going to win a 5000 dollar cash prize", even though both "may win a 5000 dollar cash prize" and "certainly going to win a 5000 dollar prize" can be true at the same time.

The spectrum of epistemic strength is semantically rich and subtle, making epistemic modality a complicated linguistic property to acquire for both native speakers and L2 learners. Hirst and Weil (1982) claimed that English speaking children start their acquisition of epistemic modality by differentiating factuals from non-factuals, and then proceed to distinguish finer semantic differences between epistemic modals. They argued further that the dichotomy between certainty/uncertainty is not available for children before the age of three (Byrnes & Duff, 1989), and probably starts to be available at the age of four or five (Hirst & Weil, 1982). Coates (1988) suggests that eight-year-old kids only have rudimentary understanding of modal meanings, and even by the age of twelve, their system of modal meanings is still not as subtle as that of the adults.

In terms of SLA research, researchers find that the first marked epistemic modality lies in the middle of the "true-probable-not true" scale, and then learners start to acquire "minus probable" and then "plus probable" (Dittmar & Terborg, 1991: 359). Although epistemic modality receives considerable attention in second language classes, even advanced second language learners of English differ significantly from native English speakers in their use of epistemic vocabulary, and such deviation is observed among students from a variety of L1 backgrounds (Chen, 2012; Carrió-Pastor, 2014; Kim & Suh, 2014; Vassileva, 2001). Problems in mastering the appropriate use of epistemic modals may even persist for L2 English learners in graduate school (Dudley-Evans,1991).

2.3 Parallels between epistemic and deontic modality

Deontic modality, also referred to as root modality (Hofmann, 1976), expresses permission and obligation, which reflects the desirability of the propositions based on the attitude

of some authority (Palmer, 2001; Verstraete, 2001; 2005). For example, “obliged to” in the sentence “You are obliged to wear sports shoes in the hall” expresses an actively enforcing attitude towards the proposition “wear sports shoes in the hall”, rather than the probability of wearing sports shoes in the hall. Given the above sentence, it is still possible (or even likely) for some people to not wear sports shoes in the hall, however, that action is considered as a violation of the rule.

Although epistemic and deontic modality are conceptually different, there are some interesting parallels between them. First of all, the same auxiliary verb in English may express either epistemic modality or deontic modality depending on the context. For example, the word “must” in “Tom is absent. He must be sick again!” expresses the speaker's belief that the probability of Tom being sick is very high, which falls into the realm of epistemic modality. By comparison, the same word “must” in the sentence “You must wear sports shoes in the hall!” expresses deontic modality. Horn (1972: 127) provides a table summary of some common English modal auxiliaries that have both the epistemic and deontic readings:

	Epistemic	Deontic
Can, Could	possibility	permission
May, Might	possibility	permission
Should, Ought	possibility	obligation
Must, Have to	certainty	obligation

Figure 2-3 Epistemic and deontic readings of some common English auxiliaries

The semantic difference between epistemic and deontic modality motivated the proposal that for the above modal auxiliaries, the different readings correspond to different underlying

syntactic structures (Perlmutter, 1968; Ross, 1969; Horn, 1972). The epistemic reading corresponds to an underlying intransitive structure, in which the embedded subject is the same subject in the surface form; while the deontic reading corresponds to an underlying transitive structure. Based on this analysis, the sentence "John may go" have two different underlying syntactic structures depending on whether it means "it is possible that John will go" or "It is ok for John to go". The syntactic structure for the first reading is shown in Figure 2-4 on top in which the event of "John goes" is the logical subject of the intransitive verb "may". By comparison, the deep structure for the second reading is transitive, in which the auxiliary "may" is equivalent to "allow" (Horn, 1972: 127).

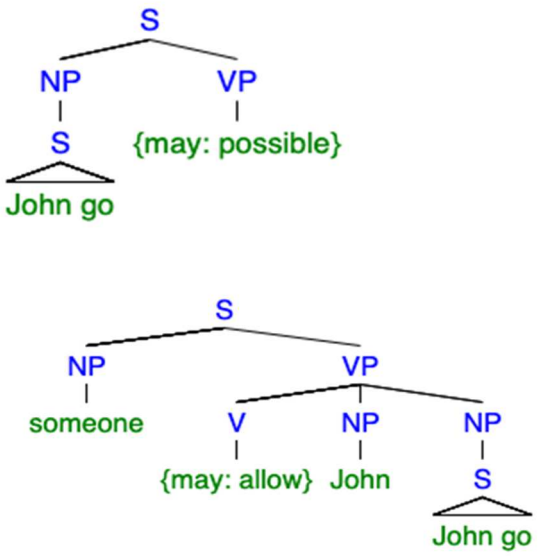


Figure 2-4 Syntactic structure corresponding to the epistemic reading (top) and deontic reading (bottom)

The second shared feature between epistemic and deontic modality is that both of them can be viewed as a quantitative scale (Horn, 1972, 1989; Levinson, 1983, 2000; see Verstraete 2005 for an alternative analysis). The semantic relationship and pragmatic implications for items on the epistemic scale can also be found among items on the deontic scale. For example,

"allowed" and "obliged" formed a pair of weaker and stronger modal values. Saying someone is obliged to do something entails that someone is also allowed to do it. On the other hand, saying someone is allowed to do something doesn't entail that someone is also obliged to; rather, pragmatically speaking, "you are allowed to" implies that "it is not the case that you are obliged to", and the underlying mechanism is the same as the one explained in the case of epistemic modals.

2.4 Nested epistemic expressions

2.4.1 Defining nested epistemic expressions

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, which is referred to in this research as "nested epistemic expressions". Halliday (1970) noted a triad pattern in the use of epistemic expressions. First, the modal auxiliary can be replaced by a non-verbal epistemic item that expresses similar degrees of probability (see 1a and 1b). Then, the verbal and non-verbal epistemic lexical items can be combined in use (1c). An example of the triad is as follows:

(1a) This gazebo may have been built by Sir Christopher Wren.

(1b) Possibly this gazebo was built by Sir Christopher Wren.

(1c) Possibly this gazebo may have been built by Sir Christopher Wren.

For Halliday (1970), sentences (1a) and (1b) express "more or less the same content" (p.328), while for sentence (1c), the two "equivalent" epistemic lexical items "reinforce each other" (p.331). In fact, for a nested epistemic expression, modals do not have to be equivalent in terms of their epistemic strength. For example, in the sentence "certainly he might have built it" (1d), the epistemic strength of "certainly" and "might" lies in the opposite polar of the scale. In this

case, the epistemic strength of the component modals pulls towards different directions, and thus these two modals are not likely to "reinforce" each other. Halliday (1970) argues that in this case, the two modals are "cumulative in meaning" (p.331), which is left vague without any further explanation.

The above two types of nested epistemic vocabulary (1c and 1d) are coined as "modally harmonic" combination (similar epistemic strength) and "modally non-harmonic" combination (contrasting epistemic strength) by Lyon (1977). According to his analysis, the modally harmonic combination is a "double realization of a single modality", and there is only one modality across the single clause. By contrast, in modally non-harmonic combination, the modal adverb and modal auxiliary "cannot but be independent; and one must be within the scope of the other" (Lyon, 1977: 808). This argument suggests different semantic structures for two types of nested modality. The non-harmonic combination specifies the sequence of embedding. For example "certainly he may have forgotten" means "it is certainly the case that he may have forgotten", rather than "it may be the case that he has certainly forgotten" (Lyon, 1977: 808). Harmonic combination, on the other hand, does not specify such sequence, because there is essentially only one modality. This idea abandons a uniform account of nested epistemic vocabulary by analyzing it differently according to different configurations of modals' relative epistemic strength.

It is important to note that the nested epistemic expressions investigated in this research were different from the case of double modals like "might could" which might be found in dialects of northern England and southern United States (Nagle, 2012). The nested epistemic expressions are different from the double modals like "might could" in certain important aspects. In terms of the syntactic category, 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 the semantic category, the nested epistemic expression consists of two epistemic modals, while double modals consist one epistemic modal and one non-epistemic (such as deontic and dynamic) modal (Nagle, 1994). Moreover, it has been argued that the double modals "might could" is one single lexical item consisting of two words, similar to a compound (Di Paolo, 1989), while nested epistemic expressions are compositional in that the two modals are believed to have different semantic scopes, and thus, the meaning of one modal should be interpreted within the meaning of the other modal (Lyons, 1977, p. 808).

2.4.2 A corpus study investigating the use of nested epistemic expressions on Twitter

In a corpus linguistics project, I surveyed the use of nested epistemic expressions in Twitter, an online microblogging and social networking platform on which users communicate with each other by posting messages known as "tweets". I built a corpus of 413986 tweets, each of which contained at least one epistemic modal. Those tweets were randomly scraped from Twitter database based on the keywords, which were some common English epistemic auxiliaries and adverbs. A summary of the keywords used for searching and the number of tweets containing each keyword is shown as below:

Keywords	Number of Tweets
Definitely	54000
Might	36000
Must	53997
Would	36000
Possibly	71999
Probably	71998
Certainly	54000
May	35992

Figure 2-5 Summary of the number of tweets in the corpus containing the keywords

The type of nested epistemic expressions I focused on were the cases containing one modal auxiliary and one modal adverb, such as “certainly may” and “might possibly”. Given that the keywords contained four modal auxiliaries and four modal adverbs, in total there were 16 different auxiliary and adverb combinations regardless of the word order, and 32 different nested epistemic expressions if the word order of the auxiliary and adverb was taken into consideration. The frequency of each of the 32 nested epistemic expressions is shown in the figure below:

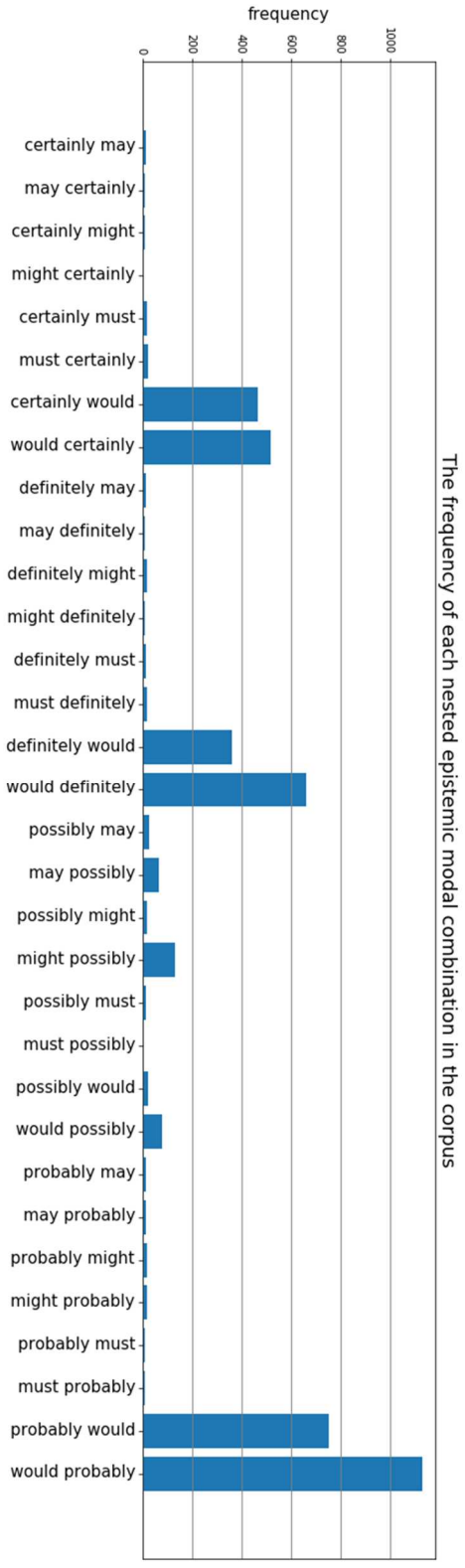


Figure 2-6 Frequency of nested epistemic expressions in the tweet corpus

It is noticeable that most of the nested expressions had a frequency lower than 200, except for those containing the modal auxiliary “would”. For example, while nested modals like “certainly may” and “certainly might” occurred less than 10 times in the tweet corpus, expressions like “would probably” occurred more than 1000 times. A closer reading of the individual tweet revealed that the word “would” often did not express epistemic modality, rather, it expressed one's willingness to do something with regard to obligation or principles, for example “I definitely wouldn't back up and go again. That would be wrong”. As mentioned in the section 2.3, the same auxiliary verb in English can express more than one type of modality depending on the conversational context. We cannot be certain about whether the modal auxiliary in a nested expression really expresses epistemic modality without examining the entire utterance and its context. However, if we suppose that the probability of an auxiliary having an epistemic reading is fixed, the relative frequency of different categories based on the auxiliary forms also reflects the relative frequency of epistemic categories. That is why, this exploratory corpus research still reveals some patterns of how people use nested epistemic expressions on the internet, though not all the modal auxiliaries in the corpus had an epistemic reading.

The first search task was to answer the question whether harmonic nested expressions were more frequent than the non-harmonic expressions. The previous literature on nested modality did not suggest that was the case, neither did it say much about the relative frequency of these two types of epistemic expressions; however, given that the two modals in the non-harmonic expression are pragmatically contradictory and thus may require greater effort to process, it is possible that non-harmonic expressions are less frequent than the harmonic expressions. Based on Halliday and Matthiessen (2004), “may”, “might” and “possible” (group 1) was lower on the scale of epistemic strength compared with “would” and “probably” (group

2) which was on the intermediate position of the scale. The word “definitely”, “must” and “certainly” were on the high end of the scale (group 3). The combination of modal auxiliary and adverb within the same group, such as “certainly must” and “possibly might”, formed harmonic nested expressions; while the combination of auxiliary and adverb across different groups, such as “definitely may” and “probably might”, created non-harmonic expressions. In total, there were 2126 harmonic expressions and 2145 non-harmonic expressions. The ratio was almost one to one, indicating that in tweets the non-harmonic nested epistemic expressions occurred as frequent as the harmonic nested epistemic expressions.

The second task focused on the non-harmonic expressions, examining the preferred word order in terms of epistemic strength. We know that in a non-harmonic nested epistemic expression, one modal was higher on the scale of epistemic strength than the other modal, and thus, for two modals in a nested expression, there are two possible ways of ordering them depending on whether the higher modal precedes or follows the lower modal. For example “certainly may” is a non-harmonic combination of high-low word order, while “may certainly” is a non-harmonic combination of low-high word order. For the 2145 cases of non-harmonic nested expressions, 932 of them were of high-low word order while 1213 of them were of low-high word order. It seems that when posting tweets, users tend to type the modal with lower epistemic strength before the modal with higher strength, but given that not all the modals in the corpus had an epistemic reading and the size of the corpus was small, the pattern mentioned above needs to be confirmed by future corpus research using a larger corpus with fine-grained semantic labels that distinguish epistemic modality from other notional categories of modality.

2.5 Processing nested epistemic expressions: two competing theories

2.5.1 The scope account

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). In this study, the scope is defined as "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.

Though it is an important concept, Hintikka pointed out that the notion of scope also leads to confusion because it has been used to express different concepts by different researchers. Hintikka observed two distinct usages of scope in linguistic literature. The first one is the notion of priority scope (or called logical scope), which is the same definition of scope used in this study. The logical scope defines the sequence in which logical operations take place. Consider the sentence "Everybody isn't happy". It has been argued that the sentence is semantically ambiguous because the relative priority of the universal quantifier and negation can be interpreted in two different ways (Coppock & Champollion, 2019). For the first interpretation, which can be formally represented as "$\forall x.\neg Happy(x)$", the universal quantifier takes a wide

scope over negation, and thus “Everybody isn’t happy” is interpreted as “for every x, it is not the case that x is happy”. The other interpretation can be formally represented as “¬∀x.Happy(x)”, in which the negation scopes over the universal quantifier. According to this analysis, the expression “Everybody isn’t happy” means “it is not the case that for every x, x is happy”.

In addition to the logical scope discussed above, the term “scope” can also be used to denote what is often called the “binding scope”, indicating a segment of a sentence where the variables are bound to a particular quantifier. For example, in sentence “If Peter owns a donkey, he beats it”, the binding scope of "the donkey" comprises the word “it” (Hintikka, 1997:519). The difference between logical scope and binding scope is illustrated by Sandu (2007: 171) using sentence “A girl smiles. She is happy”, the logic form of which can be represented as follows (square brackets and parenthesis for logical scope and binding scope respectively):

$$\exists x([G (x) \wedge S (x)] \wedge H(x))$$

According to Sandu, the logical scope of the existential quantifier (represented by the square brackets) does not include the second sentence “she is happy”. That means the syntactic operation of the clause “a girl smiles” does not extend to the second sentence “she is happy”. However, if we interpret the meaning of the sentence as “a girl smiles and that girl is happy”, the second sentence is clearly within the binding scope of the existential quantifier.

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[2] (Ladusaw, 1979: 37). On the other hand, we can define scope as a relation between the meanings of two constituents when interpreting that piece

[2] Based on Reinhart (1976), node A c(onstituent)-commands node B iff the branching node most immediately dominating A also dominates B.

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: 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" (De Swart, 1998: 177). With regard to the scope of nested modals, it has been argued that in a non-harmonic nested epistemic expression, one modal must be within the scope of the other (Lyons, 1977, p.808). Based on the syntactic structure of modal adverbs proposed by Potsdam (1998) (see Figure 2-2), the word "certainly" in sentence "he certainly may have forgotten" is the left adjunction to I', which is higher in the syntactic tree structure, c-commanding the modal auxiliary "may" (which occupies the I node). 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 the left adjunction to VP, which is c-commanded by the modal auxiliary "may" higher in the tree structure. Thus, in the sentence "he may certainly have forgotten", "certainly" is within the scope of "may".

To sum up, 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. The "modally non-harmonic" 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". The implication of the scope account is that if we change the order of the component modals in a nested expression, the meaning of the entire expression will be changed. A possible aspect of meaning that could be changed by the order of the modals is the overall strength of the expression. The perceived probability of the statement "Tom has forgotten the meeting" may be different depending on whether that statement is embedded in "Tom may certainly have forgotten the meeting" or "Tom certainly may have forgotten the meeting". The following paragraphs illustrated one possible mechanism in which the order of the nested modals influences the probability rating of the embedded statement, and this mechanism was further tested by a series of experiments reported in Chapter 3.

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 CA, 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, scope account predicts that the semantics of 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.

Based on this account, when processing non-harmonic 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 along. 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 along. 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”.

2.5.2 The good-enough processing account

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”, 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, Williams, Zacks, & Ferreira, 2006; Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 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, Hollingworth, Halliwell, & Ferreira, 2001; Christianson et al., 2006; Ferreira et al., 2001) have been 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 of the 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 percent of the time (Ferreira & Lowder, 2016). Slattery, Sturt, Christianson, Yoshida, and Ferreira (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".

2.6 Research questions

We conducted a series of six experiments investigating how comprehenders process the semantics of epistemic modals, especially the non-harmonic nested epistemic expressions. We asked the following questions which were supposed to deepen our understanding on the cognitive mechanism underlying the comprehension of epistemic expressions.

First of all, what is a native English speaker's knowledge about the strength of various epistemic modals? Existing literature suggested a three-point scale of epistemic strength with

modals like “certainly” and “must” on the high end of the scale, while modals like “may” and “possibly” are supposed to be on the lower end of the scale. This research aims to provide numerical estimates of the strength of various common epistemic modals in English, and examine to what extent the numerical estimates of the epistemic strength consistent with the proposed three-point scale.

Second, what is the compositionality of the nested epistemic expression? We studied the relation between the overall meaning of a nested epistemic expression and the meaning of the component epistemic modals, focusing on the extent to which the overall meaning of the expression is derived from the meaning of its individual components.

Third, to what extent do interlocutors process the scope of nested epistemic expressions during informal conversation? We examined the scope account and the good-enough processing account focusing on whether or not the order of the two modals has an effect on comprehenders’ interpretation of the nested expressions. If participants interpret the nested expressions differently depending on the order of the modals, it means that the scope relation has been processed. However, if the order of the modals does not influence how participants interpret the expressions, then there is no evidence showing the scope is processed. If that is the case, the mechanism proposed by the good-enough processing framework is more likely to underlie the processing of nested epistemic expressions.

Last, does the distance of the component modals influence the way nested epistemic expressions are processed? Although nested epistemic expressions are supposed to be qualitatively different from the double modal constructions (see 2.4.1), it is still possible that interlocutors interpret the component modals in a nested epistemic expression as a single lexical item if the two modals are adjacent, but not so when the two component modals are separated by

more words, such as a parenthetical element like "Bob definitely, *according to the radio*, might have hit traffic on his way home". If that is the case, we would observe different patterns in participants' interpretation of nested epistemic expressions depending on whether or not the component modals are adjacent.

References

Adger, D. (2003). *Core syntax: A minimalist approach*. Oxford: Oxford University Press.

Aĭkhenval′d, A (2014). The grammar of knowledge: a cross-linguistic view of evidentials and the expression of information source. In A. Aĭkhenval′d & R, Dixon (Eds.). *The grammar of knowledge: A cross-linguistic typology*, Oxford ; New York: Oxford University Press.1-51.

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.

Björnsson, G., & Shanklin, R. (2014). 'Must','Ought' and the Structure of Standards. In *Deontic Logic and Normative Systems*, edited by Fabrizio Cariani, Davide Grossi, Joke Meheus and Xavier Parent, Springer International Publishing. 33-48.

Byrnes, J. P., & Duff, M. A. (1989). Young children's comprehension of modal expressions. *Cognitive Development*, 4(4), 369-387

Carrió-Pastor, M. L. (2014). Cross-cultural Variation in the Use of Modal Verbs in Academic English. *SKY Journal of Linguistics*, 27, 153–166.

Chen, Z. (2012). Expression of epistemic stance in EFL Chinese university students' writing. *English Language Teaching*, 5(10), 173-179.

Chomsky, Noam. (1957). *Syntactic Structures*. The Hague: Mouton.

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

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

Cinque, G. (2006). *Restructuring and Functional Heads: The Cartography of Syntactic Structures*. Oxford: OUP.

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

Coates, J. (1988). The acquisition of the meanings of modality in children aged eight and twelve. *Journal of Child Language*, 15(2), 425-434.

Cohen, A. (2012). Generics as modals. *Recherches linguistiques de Vincennes*, 41, 63-82.

Collentine, J. (2010). The Acquisition and Teaching of the Spanish Subjunctive: An Update on Current Findings. *Hispania*, 93(1), 39-51.

Coppock, E., & Champollion, L. (2019). *Invitation to formal semantics*. Manuscript, Boston University and New York University.

De Swart, H. (1998). Licensing of negative polarity items under inverse scope. *Lingua*, 105(3-4), 175-200.

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

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

Dittmar, N., & Terborg, H. (1991). Modality and second language learning: A challenge for linguistic theory. *Linguistic theory and second language acquisition*, 347-384.

Dudley-Evans, T. (1991). Socialisation into the academic community: Linguistic and stylistic expectations of a PhD thesis as revealed by supervisor comments. *Socio-cultural issues in English for academic purposes*, 41-51.

Emonds, J. E. (1976). *A Transformational Approach to English Syntax*. New York: Academic Press.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological science*, 17(4), 311–318.

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

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

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

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

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. Academic Press.

Grice, H. P. (1975). Logic and conversation. *In Speech acts*. Brill. pp. 41-58.

Grice, H. P. (1978). Further notes on logic and conversation. In *Pragmatics.* Brill. pp. 113-127.

Halliday, M.A.K. (1970). Functional Diversity in Language as Seen from a Consideration of Modality and Mood in English. *Foundations of Language* 6: 322-61.

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

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

Hirst, W., & Weil, J. 1982. Acquisition of epistemic and deontic meanings of modals. *Journal of Child Language*, 9, 659-666.

Hofmann, T. R. (1976). Past tense replacement and the modal system. In *Notes from the linguistic underground*. Brill. pp. 85-100.

Holmes, J. (1982). Expressing Doubt and Certainty in English. *RELC Journal*, 13(2): 9-28.

Horn, L. (1972). *On the semantic properties of the logical operators in English*. (Unpublished doctoral dissertation). UCLA. Distributed by IULC, Indiana University.

Horn, L. (1989). *A Natural History of Negation*. University of Chicago Press, Chicago.

Hu, C., & Li, X. (2015). Epistemic modality in the argumentative essays of Chinese EFL learners. *English Language Teaching*, 8(6), 20-31.

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

Jacowitz, K. E., & Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21(11), 1161–1166.

Kim, C., & Suh, H. (2014). Epistemic rhetorical stance: Hedges and boosters in L1 and L2 students' English writing. *The Linguistic Association of Korea Journal*, 22(2), 61-93.

Kratzer, A. (1981). The notional category of modality. *Words, worlds, and contexts: New approaches in word semantics* (Vol. 6). Eikmeyer, H. J., & Rieser, H. (Eds.), Berlin: Walter de Gruyter. 38-74.

Kratzer, A. (2012). *Modals and conditionals: New and revised perspectives*. Oxford: Oxford University Press.

Ladusaw, W., (1979). *Polarity sensitivity as inherent scope relations*. Ph.D. thesis, University of Texas at Austin.

Levinson, S. (1983). *Pragmatics*. Cambridge university press.

Levinson, S. C. (2000). *Presumptive meanings: The theory of generalized conversational implicature*. MIT press.

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

McEnery, T., & Kifle, A. (2002). Epistemic modality in argumentative essays of second-language writers. In *Academic discourse*, J. Flowerdew (Ed.), London: Pearson Education Limited.182-195.

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

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

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

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

Palmer, F. R. (2001). *Mood and modality*. Cambridge: Cambridge University Press.

Perlmutter, D. M. (1968). *Deep and surface structure constraints in syntax* (Doctoral dissertation, Massachusetts Institute of Technology).

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

Reinhart, T. (1976). *The Syntactic Domain of Anaphora*, Doctoral dissertation, MIT, Cambridge, Massachusetts.

Ross, J. R. (1969). Auxiliaries as Main Verbs. In Todd, William (ed.), *Studies in Philosophical Linguistics,* Evanstown, IL: Great Expectations. 77-102.

Sandu, G. (2007). Two notions of scope. In A. Pietarinen (Ed.), *Game Theory and Lingustic Meaning* (*Current research in the semantics/pragmatics interface ; No. 18*). Elsevier Scientific Publ. Co. 171-182.

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.

Traugott, E. (1989). On the rise of epistemic meanings in English: An example of subjectification in semantic change. *Language*, 65(1), 31-55.

Traxler, M. J. (2011). *Introduction to psycholinguistics: Understanding language science*. John Wiley & Sons.

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

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *science*, 185(4157), 1124–1131.

Van der Auwera, J. (1996). Modality: the three-layered scalar square. *Journal of semantics*, 13(3), 181–195.

Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes*, 20(1), 83-102.

Verstraete, J. C. (2005). Scalar quantity implicatures and the interpretation of modality: Problems in the deontic domain. *Journal of pragmatics*, 37(9), 1401–1418.

Verstraete, Jean-Christophe, (2001). Subjective and objective modality. *Journal of Pragmatics* 33, 1505–1528.

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.

3 Experimental evidence for processing nested epistemic expressions[3]

3.1 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.

3.1.1 Participants

52 adult native English speakers were recruited from Amazon Mechanical Turk, a crowd-sourcing internet marketplace. All of the participants had a valid U.S. IP address, and they received monetary compensation for their participation in this study.

3.1.2 Stimuli

In this study, there were 16 experimental items which 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 the experimental items, in which, the first speaker asked a question, while the second speaker provided an answer to the question, and that answer contained an epistemic expression. As shown in Figure 3-1, the second speaker's reply fell into four experimental

[3] Section 3.1, 3.2 and 3.3 have been accepted for publication. Qiu, Z., & Ferreira, F. (n.d.). "He May Certainly Have Forgotten": The Processing of Nested Epistemic Expressions. Discourse Processes. https://doi.org/10.1080/0163853X.2022.2077064

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 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.

	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 3-1 Experiment 1: Example of an experimental item

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 of these combinations appeared in the Corpus of Contemporary American English (COCA), 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’ judgements 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 the items in this experiment can be accessed via GitHub[4].

In total, 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

[4] https://github.com/PON2020/Nested_Epistemic_Expressions_Submission

experimental item only occurred once in a list. During the experiment, all the items in the list were randomized.

3.1.3 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.

3.1.4 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 to a hundred, we transformed the rating probability into its logit following the steps below:

(1) Re-scale the rating score from 0-100 into 0-1.

(2) Re-code the re-scaled variable, for which one is coded as 0.999, while zero is coded as 0.001, following Verkuilen and Smithson (2012).

(3) Perform logit transformation on each re-coded rating score p using the equation:

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

Bayesian mixed-effects models were constructed using R package brms (Burkner, 2017) with default priors[5] 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 (see Table 3.1-1), and to Low-High condition in another model (see Table 3.1-2), 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 is available in GitHub[6]

3.1.5 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 3-2, 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

[5] 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

[6] https://github.com/PON2020/Nested_Epistemic_Expressions_Submission

"must" were used in the dialogue, participants rated the statement with the highest certainty. Moreover, all the rating scores in this study were above 50 out 100, meaning that the use of the above epistemic modals made the embedded statements sound more likely than the chance level.

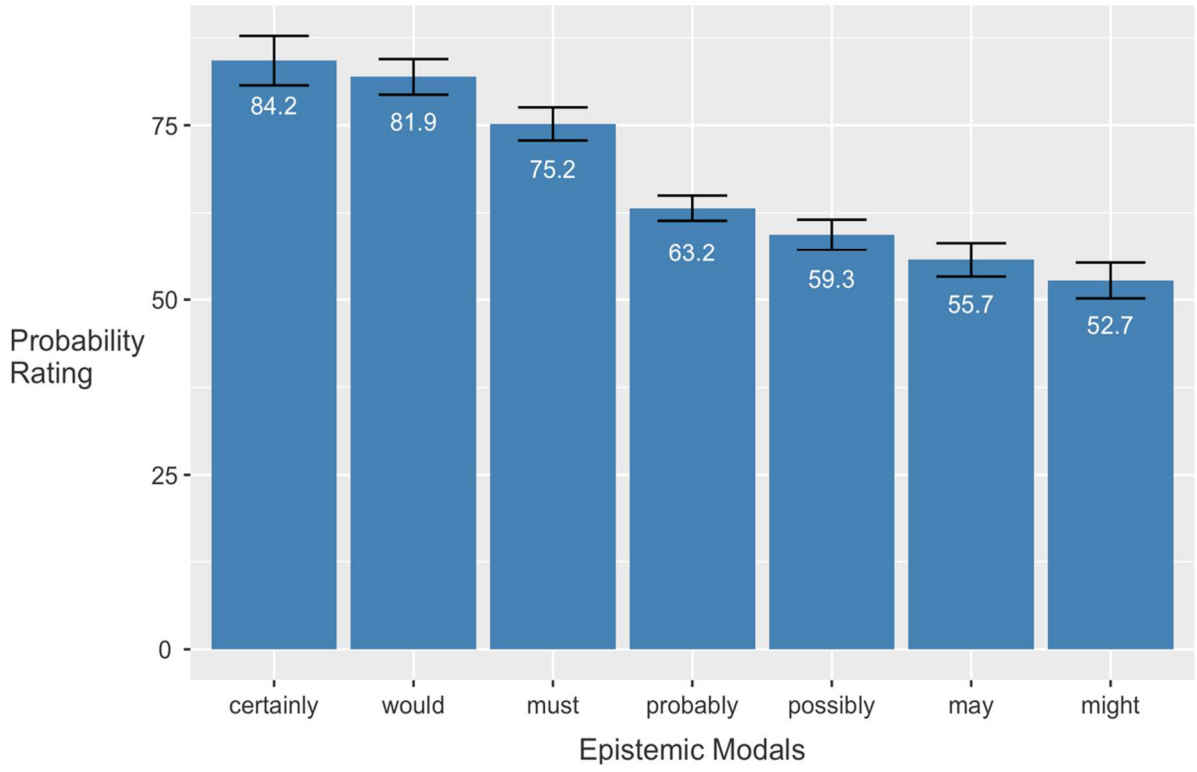


Figure 3-2 Experiment 1: Mean probability rating of different epistemic modals (with standard error)

A noticeable discrepancy between participants' rating and the prediction of linguistic theory is that whereas Halliday and Matthiessen (2004, p. 116, 623) 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 S", rather than "it is probably the case that S". Taking into consideration that Halliday and Matthiessen (2004) was not based on American English, and also, twenty 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 twenty 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 the 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 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 3-3. 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).

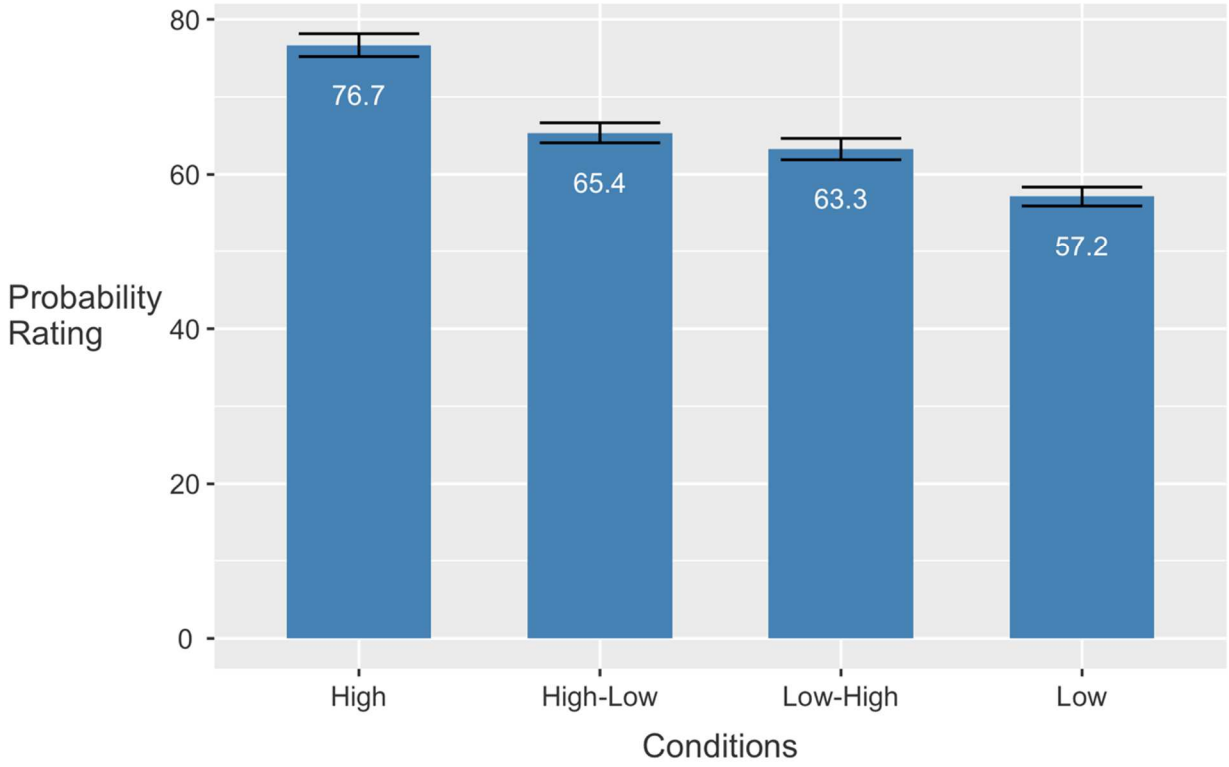


Figure 3-3 Experiment 1: Mean probability rating of different experimental conditions (with standard error)

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 Section 3.1.4. 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 (see Table 3.1-1). The same pattern was observed when the Low-High

condition was made the baseline for comparison (see Table 3.1-2). 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]).

	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 3.1-1 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	0.68	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.

Table 3.1-2 Experiment 1: Estimate of intercept and slopes from Bayesian mixed-effects model with Low-High condition as the baseline

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 (Burkner, 2017) using the Savage-Dickey density ratio method (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). When calculating BF01, three uninformative priors were selected from three normal distributions with a mean of minus 0.2 (the estimate of the coefficient from the previous Bayesian mixed-effects models, see Table 3.1-1) and a standard deviation of ten, five and one respectively. Bayes factor analyses using the priors specified above revealed moderate to strong evidence for the null hypothesis (see Table 3.1-3). 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[7].

Prior	Estimate	BF01
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.

Table 3.1-3 Experiment 1: Summary of the priors, estimates and Bayes factors in favor of the null hypothesis

[7] Fifty weakly informative priors were selected from the normal distribution with a variety of mu and sigma combinations. 90 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

3.1.6 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 being 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 of the 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. 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 Table 3.1-1 and Table 3.1-2 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 3-3), similar to the degree of probability expressed by the word "probably" (Figure 3-2). 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 has been mentioned in Section 3.1.5, 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, p. 623), "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 on 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.

3.2 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.

3.2.1 Participants

60 college students were recruited from psychology research participation system at the University of California, Davis. They participated in the study in exchange for course credits. All participants were native speakers of English and naive concerning the purposes of the experiment.

3.2.2 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 (COCA), 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 3-4.

	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 3-4 Experiment 2: Example of an experimental item

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 adding the experimental items, there were 64 items in total for each list. During the experiment, all the items in a list were randomized. The complete list of items used in this study can be found in the same GitHub inventory listed in Section 3.1.2.

3.2.3 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 (see Section 3.1.3).

3.2.4 Data analysis

This study followed the same data analysis procedure as specified in Section 3.1.4.

3.2.5 Results

The probability rating scores of individual modals in this study are shown in Figure 3-6. 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 (see Figure 3-6). 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).

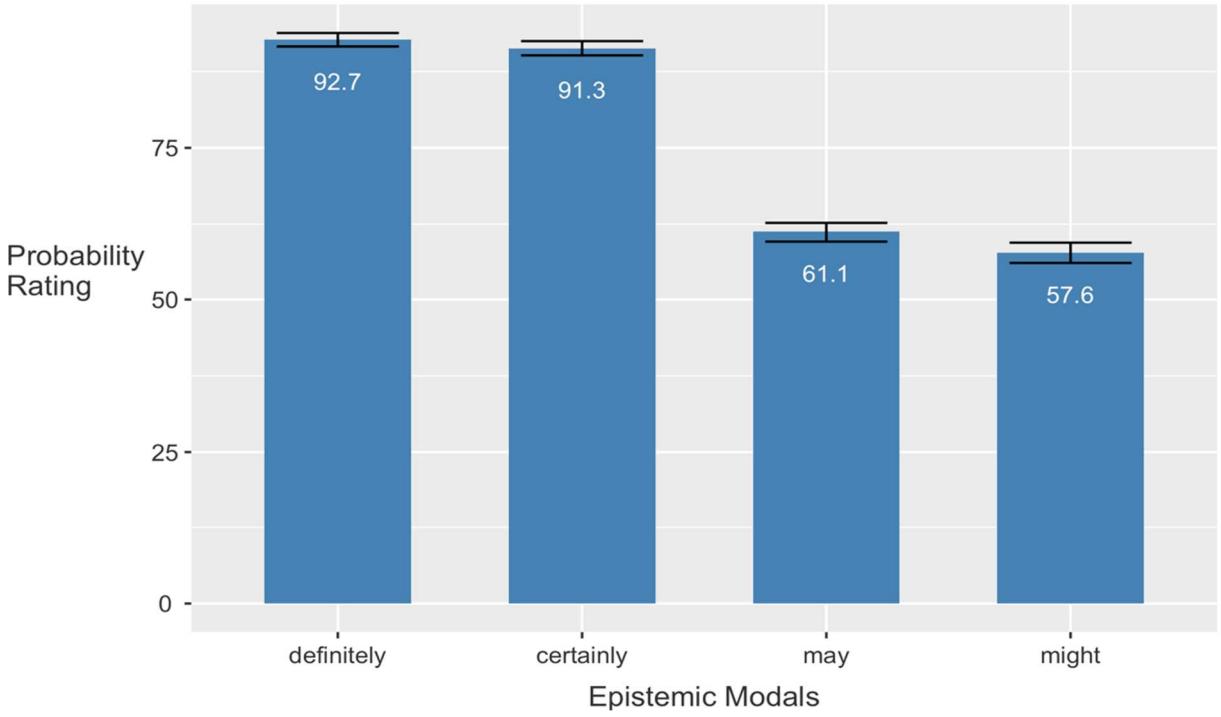


Figure 3-6 Experiment 2: Mean probability rating of different epistemic modals (with standard error)

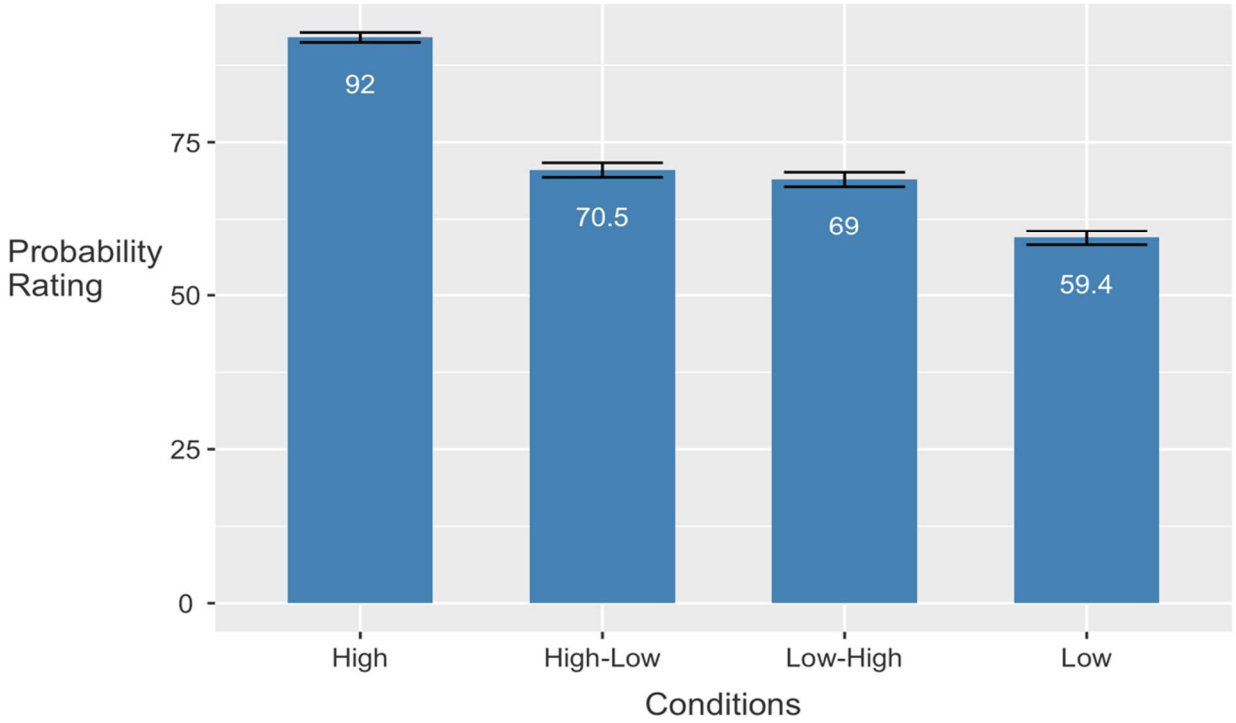


Figure 3-6 Experiment 2: Mean probability rating of different experimental conditions (with standard error)

The rating score of the nested-modal conditions (see Figure 3-6) 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 ($\beta = 2.70$, 95% CI = [2.26, 3.14]) while higher than that of the Low condition ($\beta = -0.66$, 95% CI = [-0.91, -0.40]). However, the 95% credible interval for the slope of Low-High contained the value of zero ($\beta = -0.12$, 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 (see Table 3.2-1). The same pattern was observed when the Low-High condition was made the baseline for comparison (see Table 3.2-2). The probability rating for both High condition ($\beta = 2.83$, 95% CI = [2.39, 3.28]) and Low condition ($\beta = -0.53$, 95% CI = [-0.76, -0.31]) was statistically different from the baseline, but not the High-Low condition ($\beta = 0.12$, 95% CI = [-0.14, 0.40]).

	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 3.2-1 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	1.06	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.

Table 3.2-2 Experiment 2: Estimate of intercept and slopes from Bayesian mixed-effects model with Low-High condition as the baseline

Bayes factor analyses were conducted to further evaluate the odds that there was 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 Section 3.1.5, with the only difference being the choice of priors. Three uninformative priors were selected from three normal distributions with a mean of minus 0.12 (the estimate of the coefficient from the previous Bayesian mixed-effects models, see Table 3.2-1) and a standard deviation of ten, five and one respectively. Bayes factor analyses using the priors specified above revealed moderate to strong evidence for the null hypothesis (see Table 3.2-3). 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].

[8] Fifty weakly informative priors were selected from the normal distribution with a variety of mu and sigma combinations. 96 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 that has been provided in Section 3.1.5.

Prior	Estimate	BF01
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 10-30 is considered as a strong evidence for H0, while a BF01 of 30-100 is considered as a very strong evidence for H0.

Table 3.2-3 Experiment 2: Summary of the priors, estimates and Bayes factors in favor of the null hypothesis

3.2.6 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 towards 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 towards 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, 63.3). However, in the second experiment, the patterns we found across experimental conditions were essentially the same as the 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 is possible 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 claimed 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 also 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.

3.3 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 it also provided the parser with more time to process the meaning of the first modal before encountering the second modal, and 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.

3.3.1 Participants

61 college students were recruited from the psychology research participation system at the University of California, Davis. They participated in the study in exchange for course credits. All participants were native speakers of English and naive concerning the purposes of the experiment.

3.3.2 Stimuli

This study contained 32 experimental items, each of which consisted of a written dialogue and a probability judgement 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" etc. 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 while the other expressing lower probability. All of these modal combinations had been found in Twitter or Corpus of Contemporary American English (COCA), and each of the modal combination appeared in four different experimental items. An example of the experimental item is represented in Figure 3-7.

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

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, and there were 64 items in total for each list. During the experiment, all the items in a list were randomized. The complete list of items used in this study can be found in the same GitHub inventory listed in Section 3.1.2.

3.3.3 Procedure

Participants completed the study following the procedure indicated in Section 3.2.3.

3.3.4 Data analysis

This study followed the same data analysis procedure as specified in Section 3.1.4.

3.3.5 Results

The summary of probability rating scores of individual modals in this study is shown in Figure 3-8. 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).

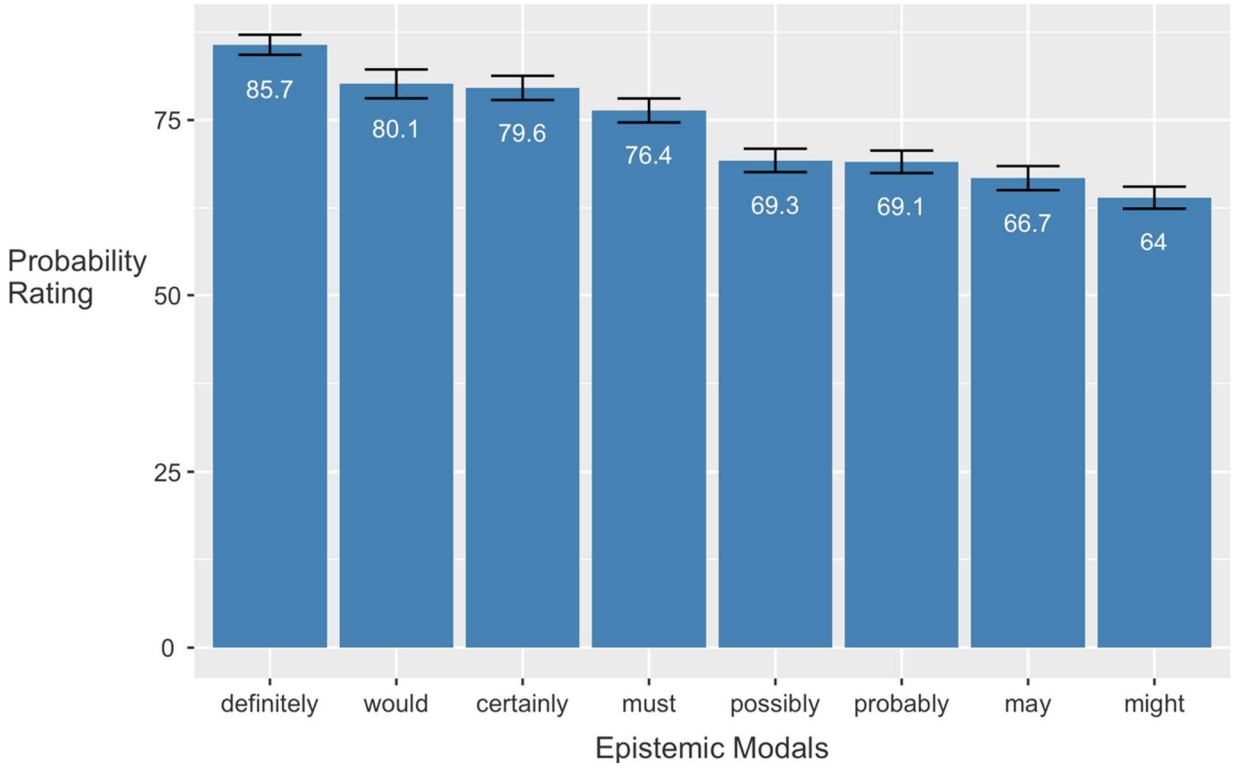


Figure 3-8 Experiment 3: Mean probability rating of different epistemic modals (with standard error)

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 3-9 A comparison between two stimuli in Experiment 1 and 3 pertaining the modal "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 3-9 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.

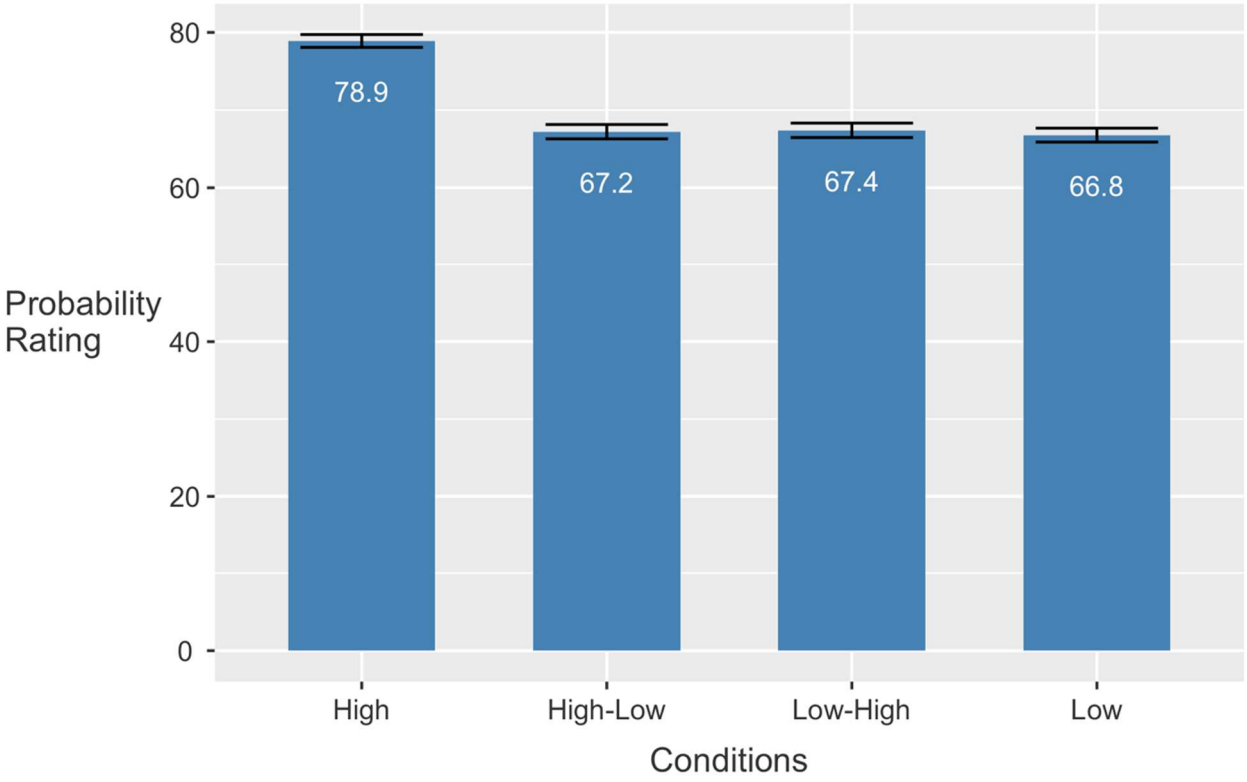


Figure 3-10 Experiment 3: Mean probability rating of different experimental conditions (with standard error)

The inflated rating scores of the low modals can also be observed across experimental conditions (see Figure 3-10). 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 the second experiment, the probability rating scores of the single low modal condition were below 60, significantly lower than the rating scores of the nested modal conditions.

	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 3.3-1 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	0.93	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.

Table 3.3-2 Experiment 3: Estimate of intercept and slopes from Bayesian mixed-effects model with Low-High condition as the baseline

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]) (see Table 3.3-1). 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 (β = -0.01, 95% CI = [-0.16, 0.14]) and the Low condition (β = -0.07, 95% CI = [-0.22, 0.08]) were essentially the same as the baseline (see Table 3.3-2).

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 Section 3.1.5, with the only difference being the choice of priors. Three uninformative priors were selected from three normal distributions with a mean of 0.01 (the estimate of the coefficient from the previous Bayesian mixed-effects models, see Table 3.3-1) and a standard deviation of ten, five and one respectively. Bayes factor analyses using the priors specified above revealed strong evidence for the null hypothesis (see Table 3.3-3). Sensitivity analyses with more informative priors confirmed that the evidence in support of the null hypothesis was not overly influenced by the prior specification[9].

Prior	Estimate	BF01
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.

Table 3.3-3 Experiment 3: Summary of the priors, estimates and Bayes factors in favor of the null hypothesis

[9] 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. The script and output of the reported Bayes factor analyses are available in GitHub following the same link that has been provided in Section 3.1.5

3.3.6 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, 2013; 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 non-epistemic (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 upon 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 quarter of worlds languages have grammatical evidentiality (Aikhenvald, 2004, p.17), especially for languages in North and South America, Caucasian languages and Tibeto-Burman languages (Song, 2018, p. 441). 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 that 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.

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. 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. This issue will be addressed in the following experiments.

3.4 Experiment 4

In this experiment, participants read English dialogues that contained epistemic expressions, and after reading each dialogue, they were instructed to rate the probability of a statement based on the information presented in the dialogue. In the dialogue, the order of the nested epistemic modals and the distance between them were manipulated to test whether readers interpret the probability of the embedded statement differently depending on the order and the distance of the modals. The scope account of the nested epistemic expressions would be supported if the difference in probability rating was observed between nested expressions with

different word orders, and we believed such a difference should be more salient when the distance between the two modals was enlarged by a parenthetical element.

3.4.1 Participants

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

3.4.2 Stimuli

This experiment contained 32 experimental items and 32 filler items, which were compiled into a questionnaire survey hosted by Qualtrics online survey platform. Each item in this experiment consisted of a written dialogue between two interlocutors, followed by a question asking participants to rate the probability of a statement based on the content of the dialogue. The format of the dialogue was the same across all 32 experimental items, in which, the first speaker asked a question, while the second speaker provided an answer to the question, and that answer contained an epistemic expression. As shown in Figure 3-11, the second speaker's reply fell into eight experimental conditions depending on the arrangement of the modals and whether or not a parenthetical element was presented.

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

only contained the modal expressing higher probability (which is “definitely” in this example) and in the Low condition, the sentence contained only the modal expressing lower probability (which is “might” in this example).

It is important to note that the label of “high” or “low” is only relative to the pair of nested modals in question. For example, “probably” is the “high” modal when paired with “might” because the degree of probability expressed by “probably” is higher than the epistemic strength of “might”; however, the same modal “probably” becomes the “low” modal when paired with the modal “must” because the epistemic strength of “must” is supposed to be higher than that of “probably”. In the example shown in Figure 3-11, the nested epistemic modals were “definitely might” (or “might definitely”), in which the modal expressing higher probability was “definitely”, while the modal expressing lower probability was “might”. Other experimental items may have a different combination of nested modals. In this study, we selected eight pairs of nested modals, which were “definitely and may”, “definitely and might”, “certainly and might”, “certainly and may”, “must and probably”, “would and possibly”, “probably and might”, and “must and possibly”. Each modal combination consisted of two epistemic modals, one expressing higher probability than the other, and each modal combination appeared in four different experimental items. All of these modal combinations had been found in Twitter or Corpus of Contemporary American English (COCA), indicating their possible occurrence in American English.

In addition to the arrangement of modals, experimental items also differed in terms of the presence of a parenthetical element indicating the source of information that the second speaker relied on when answering the first speaker’s question. In this example, the parenthetical element was the prepositional phrase (PP) “according to the radio”. Half of the experimental items had an

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

Eight lists of experimental items were created following the Latin square design so that each list contained an equal number of items in each condition, while each experimental item only occurred once in a list. In total, there were 32 experimental items in each list, and each experimental item appeared in one of the eight conditions. In addition to the experimental items, each list also included 32 filler items to hide the intended research questions from the participants. Similar to the experimental items, each filler item was also a dialogue followed by a probability judgment question. However, the filler item did not contain any epistemic modals or it only contained the negation of an epistemic modal (such "not likely"). Half of the filler items also contained a parenthetical element similar to that of the experimental items. During the experiment, all the items in the list were randomized. The inventory of all the items in this experiment can be accessed via GitHub[10].

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

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

Figure 3-11 Experiment 4: Example of an experimental item

3.4.3 Procedure

Participants used personal computers to access the web page hosting this online study, which started with demographic questions followed by the instruction and two practice trials. After the practice trials, participants hit a button to proceed. They were randomly assigned to one list of experimental items, and the first experimental item 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. Once the new experimental item revealed itself, there was no way to return to the previous item. The study ended after the participant completed all the judgment tasks in the list.

3.4.4 Data analysis

The design of this experiment was treated as two by two, in which the first factor (the arrangement of modals) had four levels and the second factor (the presence of parenthetical elements) had two levels. The probability ratings of experimental items in different conditions

were gathered and analyzed. Since the rating scores were bounded between zero to a hundred, we transformed the rating probability into its logit following the steps below:

(1) Re-scale the rating score from 0-100 into 0-1.

(2) Re-code the re-scaled variable, for which one is coded as 0.999, while zero is coded as 0.001, following Verkuilen and Smithson (2012).

(3) Perform logit transformation on each re-coded rating score p using the equation:

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

Bayesian mixed-effects models were constructed using R package brms (Burkner, 2017) with default priors to explore whether the order of the modals and the distance between them influence participants' interpretation of the nested epistemic expressions. Those mixed-effects models treated the logit of probability ratings as a function of the modal arrangement (High-Low, Low-High, Low, and High) and the presence of the parenthetical elements (Yes, No), both of which were the fixed effects of the modal. Maximal random effects structures were constructed including subject and item intercepts and slopes, following the model structure below:

rating_logit = 1+ modal arrangement * parenthesis
+ (1+ modal arrangement * parenthesis | subject)
+ (1+ modal arrangement * parenthesis | item)

The fixed effects were dummy-coded, and in order to compare each nested-model condition with the single-model conditions, we set the reference level of the first factor to the rating score of High-Low condition in one model (see Table 3.4-1), and to Low-High condition in another model (see Table 3.4-2), following the same model structure specified above. For the second

factor, the baseline condition was always the one without the parenthetical elements. The data and script used for statistical analysis of this study is available in GitHub[11].

3.4.5 Results

Participants' knowledge of the epistemic strength of each individual modal was reflected from the probability rating scores of the experimental items in single model conditions. Figure 3-12 showed the average probability rating for all the epistemic modals used in this experiment, and as we can see, participants' understanding of the degree of certainty expressed by various epistemic modals was largely consistent with the three-point scale proposed by Holmes (1982) and Halliday and Matthiessen (2004). When the epistemic modals "may", "possibly" and "might" (group 1) appeared in the dialogue, participants rated the embedded statement with the least certainty (about 65%). On the other hand, when the words "definitely", "would" and "certainly" (group 2) were used in the dialogue, participants rated the statement with the highest certainty (more than 85%). The probability rating of the model "must" (78.3%) and "probably" (70.6%) lay in between the previous two groups. All the rating scores of the epistemic modals in this study were above 60%, showing that participants interpreted the epistemic strength of all these modals as above the chance level, and thus, all the modals used in this experiment were on the scale of positive epistemic strength.

[11] https://github.com/PON2020/Nested_Epistemic_Expressions_Further_Inquiry/tree/main/Data_Analysis

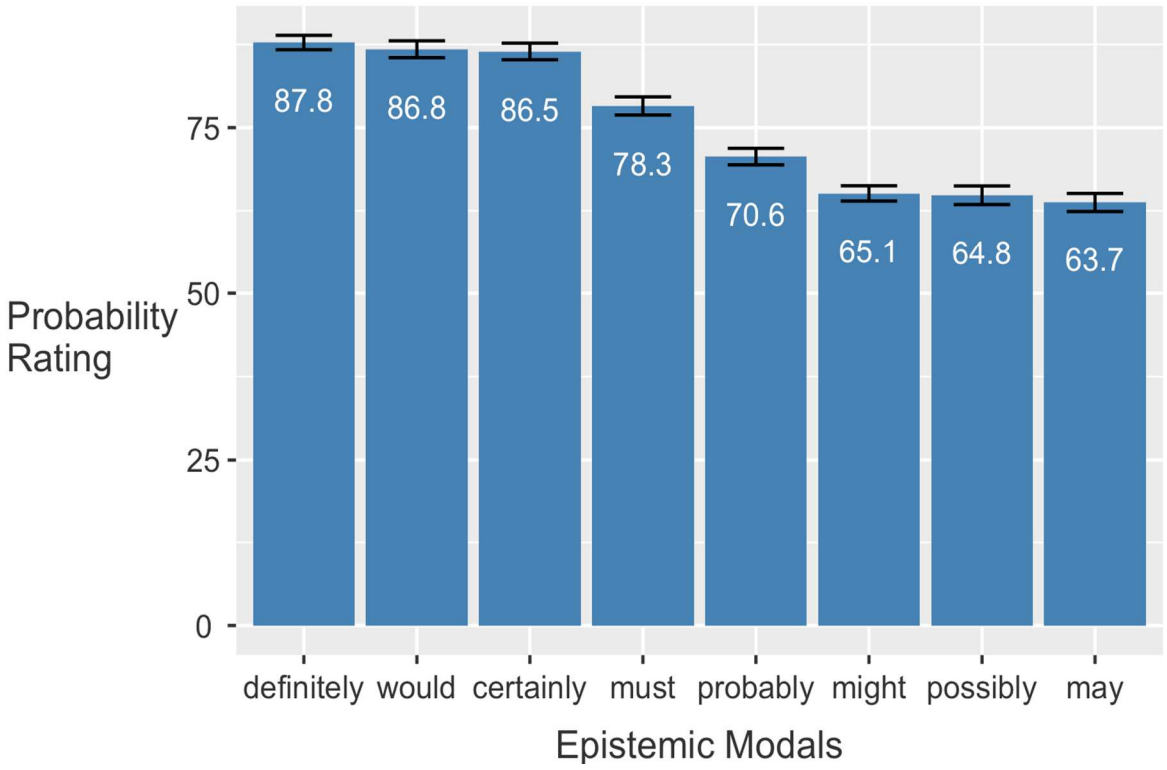


Figure 3-12 Experiment 4: Mean probability rating of different epistemic modals (with standard error)

The probability rating for items in different experimental conditions is shown in Figure 3-13. We found that items containing a single modal of higher epistemic strength received the highest probability rating, while items containing a single modal of lower epistemic strength received the lowest probability rating. This pattern was consistently observed regardless of whether or not a parenthetical element was presented in the dialogue. For this experiment, the statements in High conditions were rating above 80 out of 100, while the statements in the Low conditions were rated below 70. The rating score of the nested-modal conditions lay in between the rating scores of single-modal conditions, with High-Low condition rated slightly lower than the Low-High condition.

Figure 3-13 Experiment 4: Mean probability rating of different experimental conditions (with standard error)

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 Section 3.4.4. When High-Low condition was the baseline for comparison (see Table 3.4-1), it was rated lower than the High condition ($\beta = 1.58$, 95% CI = [1.09, 2.08]) but higher than the Low condition ($\beta = -0.35$, 95% CI = [-0.63, -0.08]). Importantly, the 95% credible interval for the slope of Low-High contained the value of zero ($\beta = 0.17$, 95% CI = [-0.13, 0.46]), meaning it is likely that there was no statistical difference between the High-Low condition and Low-High condition when the parenthetical element was not present. The same pattern was observed when the Low-High

condition was made the baseline for comparison (see Table 3.4-2). The probability rating for both the High condition (β = 1.42, 95% CI = [0.92, 1.92]) and the Low condition (β = -0.52, 95% CI = [-0.79, -0.25]) was statistically different from the baseline, but not the High-Low condition (β = -0.17, 95% CI = [-0.45, 0.12]).

In terms of the presence of the parenthetical element in the nested modal conditions, neither of the two models in Table 3.4-1 and Table 3.4-2 found the main effect of parenthetical elements (β = -0.24, 95% CI = [-0.04, 0.52] for one model and β = 0, 95% CI = [-0.27, 0.28] for the other). This means for a nested expression with two epistemic modals, adding a parenthetical element in between the two modals did not change participants' rating of the embedded statement in that expression. Moreover, neither of the two models revealed any interaction between the order of the nested modals and the presence of the parenthetical element (β = -0.24, 95% CI = [-0.66, 0.18] for one model, and β = 0.24, 95% CI = [-0.19, 0.65] for the other), showing that the lack of effect of the parenthetical element was the same for both High-Low and Low-High condition.

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.06	0.14	0.78	1.34
Low-High	0.17	0.15	-0.13	0.46
High	1.58	0.25	1.09	2.08
Low	-0.35	0.14	-0.63	-0.08
With-Parenthesis	0.24	0.14	-0.04	0.52
Low-High: Parenthesis	-0.24	0.21	-0.66	0.18

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

Table 3.4-1 Experiment 4: Estimate of main effect intercept and slopes from Bayesian mixed-effects model with High-Low condition without parenthetical elements as the baseline

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.22	0.14	0.94	1.51
High-Low	-0.17	0.15	-0.45	0.12
High	1.42	0.25	0.92	1.92
Low	-0.52	0.14	-0.79	-0.25
With-Parenthesis	0.00	0.14	-0.27	0.28
High-Low: Parenthesis	0.24	0.22	-0.19	0.65

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

Table 3.4-2 Experiment 4: Estimate of main effect intercept and slopes from Bayesian mixed-effects model with Low-High condition without parenthetical elements as the baseline

To further explore possible interactions between the presence of parenthetical elements and the strength of a single epistemic modal, we built two Bayesian mixed-effects models to analyze the data in single modal conditions. These two models followed the same structure as the previous models, treating the logit of probability ratings as a function of modal arrangement and the presence of the parenthetical elements. In one model, the baseline for comparison was the High condition without parenthetical elements (Table 3.4-3), and for the other model, the baseline for comparison was the Low condition without parenthetical elements (Table 3.4-4).

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	2.65	0.23	2.19	3.11
Low	-1.94	0.24	-2.42	-1.45
With-Parenthesis	-0.51	0.28	-1.06	0.04
Low:Parenthesis	0.77	0.31	0.18	1.37

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

Table 3.4-3 Experiment 4: Estimates of fixed effects for data in single modal conditions (baseline: High condition without parenthesis)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.71	0.11	0.50	0.92
High	1.94	0.25	1.45	2.44
With-Parenthesis	0.26	0.11	0.04	0.48
High:Parenthesis	-0.76	0.32	-1.40	-0.14

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

Table 3.4-4 Experiment 4: Estimates of fixed effects for data in single modal conditions (baseline: Low condition without parenthesis)

When the High modal condition without parenthetical element was the baseline (see Table 3.4-3), statements embedded in Low modals were rated lower than the baseline (β = -1.94, 95% CI = [-2.42, -1.45]). Although for the High modal condition, adding a parenthetical element decreased the rating score in a way that was not statistically meaningful (β = -0.51, 95% CI = [-1.06, 0.04]), if the parenthetical element was added to the Low modal condition, it increased the logit of probability rating by an extra 0.77 (95% CI = [0.18, 1.37]) . When the Low modal condition without parenthetical element was the baseline (see Table 3.4-4), adding a parenthetical element in general would increase the rating score (β = 0.26, 95% CI = [0.04, 0.48]); however, if the parenthetical element was added to the High modal condition, it decreased the logit of probability rating by an extra 0.76 (95% CI = [-1.40, -0.14]).

3.4.6 Discussion

When the participants read a statement embedded in the scope of an epistemic modal, the probability rating of that statement reflected participants' knowledge about the strength of that modal. The findings of this experiment showed that participants were sensitive to the epistemic strength of the modals in the dialogue, and thus they made reasonable inferences about the

probability of the embedded statement. For instance, the sentence "Tom certainly has forgotten." and "Tom may have forgotten." both indicate the probability of "Tom has forgotten". This means the same proposition "Tom has forgotten" is embedded within the scope of the epistemic modal in both of the sentences. The difference between these two sentences lies in the semantics of the modal. While "certainly" expresses high probability, the word "may" expresses low probability. Participants were able to calculate the probability of the statement by applying the semantics of the epistemic modal to the statement within its scope, and thus in this experiment, the embedded statements in the Low condition were rated noticeably lower than the same statements in the High condition (β = -1.94, 95% CI = [-2.42, -1.45], see Table 3.4-3).

The epistemic strength of each modal, as reflected from the probability rating, was in general consistent with the proposed scale of epistemic strength (Halliday & Matthiessen, 2004; Holmes, 1982). The only difference between participants' rating and the proposed three-point scale was the rated probability of modal "would" and "must". While Halliday and Matthiessen (2004, p. 116, 623) proposed that the word "would" expressed medium probability and "must" expressed high probability, the reversed pattern was observed among participants' rating in this study. In general, participants interpreted "would S" (the letter S stands for an embedded statement) as an expression similar to "it is definitely the case that S" or "it is certainly the case that S", rather than "it is probably the case that S"; on the other hand, they rated "must S" as an expression similar to "it is probably the case that S" which had lower strength than what Halliday and Matthiessen (2004) has argued. It is important to note that Halliday and Matthiessen (2004) was published twenty years ago and was not based on American English, and thus it is not unexpected that some epistemic modals is interpreted slightly different from what was believed twenty years ago. Given that the focus of this study was on the processing of nested epistemic

expressions, namely the High-Low and Low-High conditions, our experimental manipulation would make sense as long as the modals within each pair of nested combination expressed different degree of probability. As mentioned in the previous paragraph, the probability rating of single-modal conditions showed that for all the modal combinations of this experiment, the High modal indeed expressed higher probability than the Low modal, confirming that the nested epistemic expressions in this study were non-harmonic combinations.

As to the nested epistemic expressions, the scope account predicts that if the order of the two component modals has changed, the overall meaning of the nested expression will change accordingly. Thus, a difference in probability rating between the High-Low and Low-High conditions would be a minimal supporting evidence for the scope account. In this experiment, we found that when the component modals were adjacent to each other, statements embedded in the nested-modal conditions were rated higher than the statement embedded in the single Low modal condition but lower than statement embedded in High modal condition. This pattern was observed regardless of whether the High-Low (see Table 3.4-1) or Low-High (see Table 3.4-2) was the baseline for comparison. This suggested that when the component modals are adjacent, the epistemic strength of the nested-expression lies in between the individual epistemic strength of the component modals. Moreover, there was no difference between High-Low and Low-High conditions in probability rating. For nested epistemic expressions, changing the word order of the two modals did not change the way people interpreted the probability of the embedded statement. Thus, we did not find supporting evidence for the scope account of the nested epistemic expressions in this experiment.

The lack of ordering effect held true regardless of the presence of the parenthetical elements. This means sentences like “Bob certainly, the neighbor suggested, may be in the candy

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

Although we didn't observe an interaction between parenthetical elements and word order in the nested-modal conditions, the single-modal data suggested an interaction effect between the strength of individual modal and the presence of parenthetical elements. When a statement was embedded within a single modal of lower epistemic strength, adding a parenthetical element would increase the probability of the statement; by contrast, when a statement was embedded within a single modal of higher epistemic strength, adding a parenthetical element would decrease the probability of the statement. The parenthetical elements used in this experiment were expressions indicating the source of information upon which the second speaker's statement was based, such as "based on what I heard", and "according to the forecast", etc. These parenthetical elements belong to the notional category of evidentials, which are expressions or grammatical markers that "indicate something about the source of the information in the proposition" (Bybee, 1985, p. 184). In general, adding the source of information changes the credibility of the message as a whole. For example, compare "Bob,

according to the radio, might have hit traffic on his way home" with "Bob might have hit traffic on his way home". If we believe that the radio is a reliable information source, adding the parenthetical evidential should make the epistemic expression "Bob might have hit traffic on his way home" more credible. What we found in this experiment was that for stimuli in the single Low modal condition, the perceived probability of the embedded statement increased when the information source was added to the epistemic expression, while for stimuli in the High modal condition, the perceived probability of the embedded statement decreased when the information source was added to the epistemic expression.

It seems that the parenthetical evidential functions like a hedge, pulling the epistemic strength of the modals from more extreme position to the middle ground. When a statement is embedded in the scope of an epistemic modal expressing very high probability, adding a parenthetical evidential softens the strength of the modal. Compare "The new secretary *certainly* has forgotten the meeting", with "The new secretary, as my previous experience suggests, *certainly* has forgotten the meeting". The findings of this experiment suggested that participants rated the probability of the new secretary forgetting the meeting differently depending on whether they read the first sentence or the second sentence. The first sentence without a parenthetical element was rated higher in strength than the one with a parenthetical element. On the other hand, if a statement was embedded in an epistemic modal expressing very low probability like "The new secretary *might* have forgotten the meeting", adding the same parenthetical element increased the overall epistemic strength (like "The new secretary, as my previous experience suggests, *might* have forgotten the meeting"). Reinhart (1983) suggested that the meaning of some evidentials is the same as that of epistemic modals, and if that is true, the expression containing an epistemic modal and an evidential expression can be regarded as a

nested epistemic expression. We will dive deeper into this proposal in the General Discussion session.

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

3.5 Experiment 5

This experiment replicated Experiment 4 using a modified paradigm that resembled daily informal conversation. It created a more natural context to investigate the processing of nested epistemic expression focusing on the effect of word order and the presence of parenthetical elements.

3.5.1 Participants

109 college students were recruited from psychology research participation system at the University of California, Davis. They participated in the study in exchange for course credits. All participants were native speakers of English and naive concerning the purposes of the

experiment. Participants were excluded if their accuracy rate for attention-checking items was lower than 90%. As a result, the data from 80 out of 109 participants were included in the data analyses.

3.5.2 Stimuli

All the stimuli in the previous experiment described in Section 3.4.2 were read and recorded by two native speakers of American English. For each dialogue, the line of the first interlocutor was read by the female speaker, while the line of the second interlocutor was read by the male speaker. The recordings were edited using the software program Praat (Boersma, 2011), which standardized the intensity of all the stimuli to 70 dB. Moreover, in each experimental item, the acoustic properties of the same epistemic modal maintained the same across all conditions, eliminating potential confounding effects of the acoustics properties of the epistemic expressions on participants' rating. The question following the dialogue was not included in the recording, and was presented visually.

The filler items in this study were also adopted and recorded from those in Experiment 4. We selected 13 filler items as the attention-checking items to detect possible disengagement from the tasks during the experiment. Those checking items were also dialogues followed by probability judgment questions, while the probability of the statements was either extremely possible or extremely impossible, and thus, if a participant's rating was in the opposite direction, that trail would be marked as incorrect. An example of the checking items is as below:

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

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

3.5.3 Procedure

Participants used personal computers to access the web page hosting this online study, which started with demographic questions followed by the instruction and three practice trials. The practice trails mimicked the format of the experimental items in which the dialogues were presented acoustically and the questions were displayed visually. Participants were instructed to press the play button to hear the recorded dialogue, and after that, read the question and rate the probability of a statement in the way described in Section 3.4.3. The recordings in this experiment can be played only once. After the practice trials, participants hit a button to proceed. They were randomly assigned to one list of experimental items, and the first experimental item in the list was presented on the computer monitor screen. Participants listened to the dialogue and provided probability rating the same way as they did in the practice trails. After that, they hit the proceed button to reveal the next experimental item. Once the new experimental item revealed itself, there was no way to return to the previous item. The study ended after the participant completed all the judgment tasks in the list.

3.5.4 Data analysis

This study followed the same data analysis procedure as specified in Section 3.4.4.

3.5.5 Results

The probability rating of individual modals are shown in Figure 3-14. Similar to Experiment 4, the degree of probability expressed by different epistemic modals revealed a scale from high probability (such as "definitely" and "certainly") to intermediate probability (such as "probably") and to low probability (such as "might" and "may"). Though the general profile of the scale obtained in this experiment remained largely the same as that of the previous experiment, a couple of differences can be found in the rating of some specific epistemic modals. Compared with the Experiment 4, the rating of the modal "would" decreased from 86.8 out of 100 to 80.4 out of 100, moving away from the high end of the scale, but closer to the intermediate position of the scale. At the low end of the scale, the rating of modal "may" and "might" decreased slightly from about 64 out of 100 to about 60 out of 100.

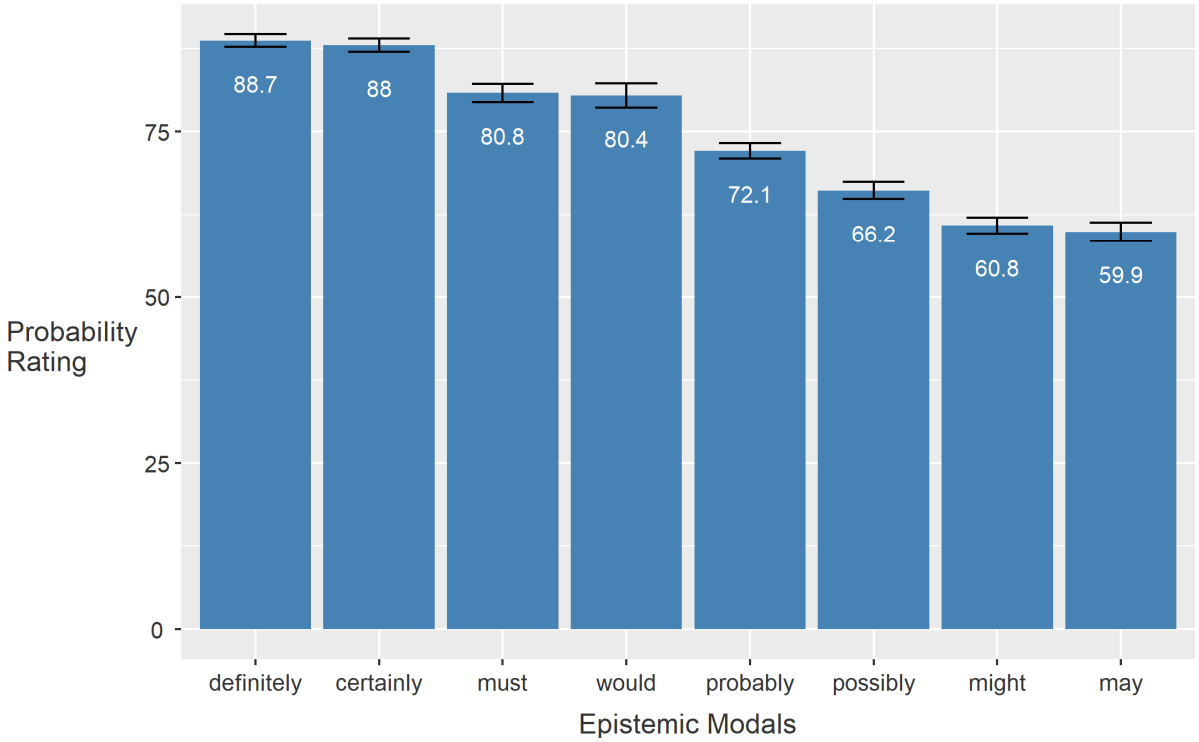


Figure 3-14 Experiment 5: Mean probability rating of different epistemic modals (with standard error)

The probability rating score across different experimental conditions echoed the patterns we found in the previous experiment. In general, statements embedded within the scope of a single High modal received the highest rating (83.8 and 82.7) while the statements embedded within the scope of a single Low modal received the lowest rating (60.3 and 66.7), especially when the parenthetical element was not present. In general, statements in the nested-modal conditions were rated in between the ratings of the High and Low conditions, with the Low-High condition being related slightly higher than the High-Low condition. Two Bayesian mixed-effects models were constructed estimating the logit of the probability rating score across conditions following the steps illustrated in Section 3.4.4. When the High-Low condition without the presence of the parenthetical element was the baseline for comparison (see Table 3.5-1), it was rated lower than the High condition ($\beta = 1.81$, 95% CI = [1.37, 2.26]) but higher than the Low condition ($\beta = -0.33$, 95% CI = [-0.53, -0.14]). Similarly, when the Low-High condition without a parenthetical element was the baseline for comparison (see Table 3.5-2), it was rated lower than the High condition ($\beta = 1.76$, 95% CI = [1.32, 2.20]) but higher than the Low condition ($\beta = -0.38$, 95% CI = [-0.58, -0.19]).

A closer look at the nested-modal conditions showed that the order of the two modals did not influence the rating of the embedded statement. When High-Low condition without parenthetical element was the baseline for comparison, the 95% credible interval for the slope of Low-High condition included the value of zero ($\beta = 0.05$, 95% CI = [-0.15, 0.26]), meaning it is likely that there was no statistical difference between the High-Low condition and Low-High condition when the parenthetical element was not present. When the Low-High condition without parenthetical elements was the baseline for comparison, the 95% credible interval for the slope of High-Low also included zero ($\beta = -0.05$, 95% CI = [-0.24, 0.15]).

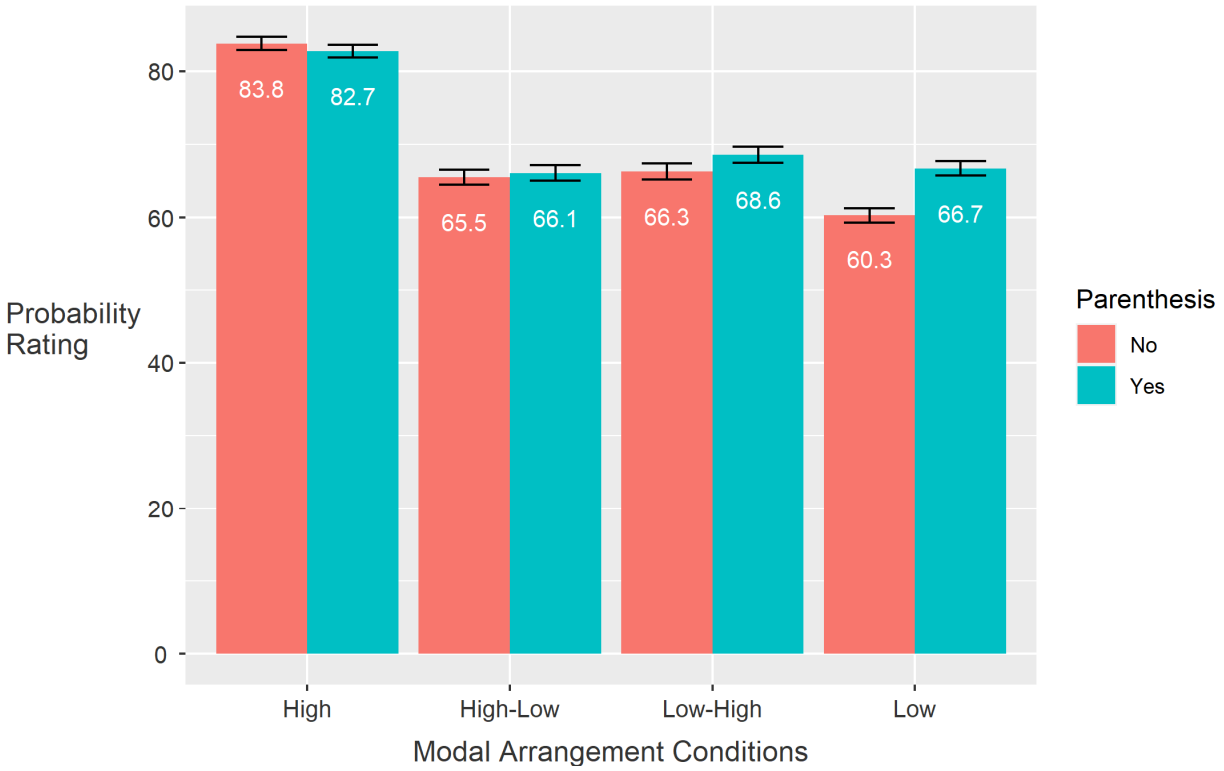


Figure 3-15 Experiment 5: Mean probability rating of different experimental conditions (with standard error)

In terms of the presence of the parenthetical element in the nested modal conditions, neither of the two models in Table 3.5-1 and Table 3.5-2 found the main effect of parenthetical elements (β = -0.03, 95% CI = [-0.25, 0.19] for one model and β = 0.1, 95% CI = [-0.11, 0.31] for the other). This means for a nested expression with two epistemic modals, adding a parenthetical element in between the two modals did not change participants' rating of the embedded statement in that expression. Moreover, neither of the two models revealed any interaction between the order of the nested modals and the presence of the parenthetical element (β = 0.13, 95% CI = [-0.19, 0.46] for one model, and β = -0.13, 95% CI = [-0.45, 0.19] for the other), showing that the lack of effect of the parenthetical element was the same for both High-

Low and Low-High condition. This pattern was essentially the same as what had been observed in Experiment 4.

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.86	0.11	0.64	1.08
Low-High	0.05	0.10	-0.15	0.26
High	1.81	0.23	1.37	2.26
Low	-0.33	0.10	-0.53	-0.14
With-Parenthesis	-0.03	0.11	-0.25	0.19
Low-High:Parenthesis	0.13	0.16	-0.19	0.46

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

Table 3.5-1 Experiment 5: Estimate of main effect intercept and slopes from Bayesian mixed-effects model with High-Low condition without parenthetical elements as the baseline

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.91	0.11	0.69	1.13
High-Low	-0.05	0.10	-0.24	0.15
High	1.76	0.23	1.32	2.20
Low	-0.38	0.10	-0.58	-0.19
With-Parenthesis	0.10	0.11	-0.11	0.31
High-Low:Parenthesis	-0.13	0.16	-0.45	0.19

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

Table 3.5-2 Experiment 5: Estimate of main effect intercept and slopes from Bayesian mixed-effects model with Low-High condition without parenthetical elements as the baseline

To evaluate the interaction effect observed in Experiment 4 between the epistemic strength of a single modal and the presence of a parenthetical element, we built two Bayesian mixed-effects models following the same procedure listed in Section 3.4.5 to analyze the rating

score in single-modal conditions as a function of the epistemic strength and the presence of a parenthetical element. Similar to what we found in Experiment 4, when the High modal condition without parenthetical element was the baseline (see Table 3.5-3), statements embedded within the scope of Low modals were rated lower than the baseline (β = -2.15, 95% CI = [-2.58, -1.72]). Adding a parenthetical element decreased the rating score in a way that was not statistically meaningful (β = -0.47, 95% CI = [-0.97, 0.03]), however, if the parenthetical element was added to the Low modal condition, it increased the logit of probability rating by an extra 0.9 (95% CI = [0.35, 1.48]). By comparison, when the Low modal condition without parenthetical element was the baseline (see Table 3.5-4), adding a parenthetical element in general would increase the probability rating score of the embedded statement (β = 0.43, 95% CI = [0.20, 0.67]); however, if the parenthetical element was added to the High modal condition, it decreased the logit of probability rating by an extra 0.89 (95% CI = [-1.46, -0.31]). The interaction effect between epistemic strength and the presence of the parenthetical element observed in Experiment 4 was replicated.

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	2.67	0.23	2.22	3.12
Low	-2.15	0.22	-2.58	-1.72
With-Parenthesis	-0.47	0.25	-0.97	0.03
Low:Parenthesis	0.90	0.28	0.35	1.48

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

Table 3.5-3 Experiment 5: Estimates of fixed effects for data in single modal conditions (baseline: High condition without parenthesis)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.52	0.11	0.32	0.73
High	2.14	0.23	1.69	2.59
With-Parenthesis	0.43	0.12	0.20	0.67
High:Parenthesis	-0.89	0.29	-1.46	-0.31

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

Table 3.5-4 Experiment 5: Estimates of fixed effects for data in single modal conditions (baseline: Low condition without parenthesis)

3.5.6 Discussion

In this experiment, we examined the findings of Experiment 4 using a modified paradigm that closely mimicked informal conversation. Instead of reading the script of a dialogue containing epistemic expressions, participants in this experiment heard the dialogue unfolding the same way as they overheard a conversation between two interlocutors in daily situation. This created a more natural context for participants to process the nested epistemic expression in the dialogue. The major findings of the previous experiments were successfully replicated.

First of all, though small fluctuations were observed in the probability rating of some modals, in general, the rating of various epistemic modals revealed a three-point scale from high probability to intermediate probability to low probability. The difference in rating between the modals expressing higher probability (High condition) and lower probability (Low condition) was statistically meaningful, and such a difference was apparent even when a parenthetical element was present. Although adding a parenthetical element would not make the rating of the Low modal condition higher than that of the High modal condition, it boosted the probability

rating of the statement within the scope of a single Low modal to a statistically significant extent. Interestingly, the effect of adding a parenthetical element was quite the opposite for the embedded statements in the High modal condition. Both this experiment and the previous experiment showed that the presence of a parenthetical element slightly decreased the rating of the embedded statement in the High condition.

Compared with the single-modal conditions, the nested-modal conditions showed almost no effect of the parenthetical element. In general, the mean rating score of the High-Low condition was slightly lower than that of the Low-High condition, regardless of the presence of the parenthetical element. However, such a small difference was not statistically meaningful. Moreover, for the nested-modal conditions, the slope for the parenthetical element had a credibility interval including zero, regardless of which baseline had been chosen for comparison. This suggested that regardless of the presence of the parenthetical element, changing the order of the modals in a nested expression would not change the perceived probability of the embedded statement.

The findings of this experiment again did not support the scope account, which predicted that the meaning of the nested expressions depends on the order of the two component modals, and when the two modals in a nested expression switch their positions, the meaning of the expression would have changed. We used the strength of the epistemic expression as a proxy for its meaning, and found that for the nested epistemic expressions, changing the order of the component modals did not change how people interpreted the strength of the expression. It is likely that when interlocutors encountered an epistemic expression containing "non-harmonic" modals, they would interpret such expression as an indicator of medium degree of uncertainly in between the epistemic strength of the component modals. For example, given a statement "S",

the expression of both "certainly may S" and "may certainly S" mean that the speaker conveys uncertainty over the probability of "S", the strength of which is estimated to be higher than "may" but lower than "certainly". This interpretation is consistent with the good-enough processing framework (Ferreira & Lowder, 2016).

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

experiment reported in the next section adopted this approach to explore the processing of nested epistemic expressions focusing on semantic aspects other than the overall epistemic strength.

3.6 Experiment 6

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

3.6.1 Participants

128 college students were recruited from the psychology research participation system at the University of California, Davis. They participated in the study in exchange for course credits. All participants were native speakers of English and naive concerning the purposes of the experiment. Following the pre-registered criteria of data inclusion (ASPREDICTED #51982 available via https://aspredicted.org/blind.php?x=FNY_YLK), participants who failed to achieve an accuracy rate above 90% for attention-checking filler items were excluded from analysis. As a result, data of 96 out of 128 participants were analyzed.

3.6.2 Stimuli

This study contained 48 experimental items. Each of the experimental items had a prompt describing a decision making context in which a group of 10 people voted to make a collective decision. The decision was described by a statement that contained epistemic modal(s), which fell into four different experimental conditions.

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

Figure 3-16 Experiment 6: Example of an experiment item.

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

As shown in Figure 3-16, the decision making context described in the prompt of this experimental item was 10 ladies voting for the best pie in a competition. The agreement that the committee reached was described by a statement which was manipulated in terms of the

epistemic modal(s) it had. In the High-Low condition, the epistemic modal expressing higher probability (which is the word "certainly") preceded the modal expressing lower probability (which is the word "may"). In the Low-High condition, the epistemic modal expressing higher probability followed modal expressing lower probability. Those two conditions were the nested-modal conditions since more than one epistemic modals occurred in the same clause. By comparison, the sentence in the single-modal conditions only contained one epistemic modal expressing either higher probability or lower probability.

After reading the prompt, participants were instructed to select from two options the one that was most consistent with the statement in the prompt. Those options described voting scenarios in which the number of the voters and/or the proposition they agreed upon differed slightly. Those options were constructed based on the mental committee framework (Moss, 2015) to reflect the semantics of the epistemic expression in the statement. For an experimental item in the single-modal condition, one of the options was supposed to be semantically equivalent to the prompt in the High condition, while the other option was supposed to be semantically equivalent to the Low condition. In the example shown in Figure 3-16, the prompt of the Low condition, which said " Mrs. Gherkin's blueberry pie may win the purple ribbon", is most consistent with the voting scenario in which six ladies on the church committee voted for Mrs. Gherkin's blueberry pie. By comparison, the prompt of the High condition, " Mrs. Gherkin's blueberry pie is certainly going to win the purple ribbon", corresponded to the option saying "8 ladies on the church committee are going to vote for Mrs. Gherkin's blueberry pie". For the two options in the single-modal conditions, the number of the voter reflected the epistemic strength of the modal in the prompt. For instance, the epistemic strength of the modal "certainly", as reflected from the rating of Experiment 4 and 5, was about 80 out of 100, the ratio

of which is equivalent to eight out of ten committee members. Similarly, the epistemic strength of the modal "may" on average was about 60 out of 100, which was represented by six out of ten committee members in the option. A list of epistemic modals and its corresponding interpretation based on the framework of mental committee is shown in Figure 3-17.

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

Figure 3-17 Epistemic modals and its interpretation based on the framework of mental committee

The two options for the nested-modal conditions were more complicated, which contained the number of the voters and an embedded epistemic modal. For an experimental item in the nested-modal condition, one of the options was supposed to be semantically equivalent to the prompt in the High-Low condition, while the other option was supposed to be semantically equivalent to the prompt in the Low-High condition. According to the scope account, the meaning of the sentence in High-Low prompt, "Mrs. Gherkin's blueberry pie certainly may win the purple ribbon", means "it is certainly the case that Mrs. Gherkin's blueberry pie may win the purple ribbon". This can be represented as eight out of ten ladies agree that they may vote for Mrs. Gherkin's blueberry pie based on the mental committee framework. Similarly, the prompt in the Low-High condition can be represented as "6 ladies on the church committee agree they

are certainly going to vote for Mrs. Gherkin's blueberry pie". Participants were asked to pick up the option that was most consistent with the prompt, and if participants constantly matched the prompt to the target voting scenarios predicted by the scope account, and also, the nested expressions in different word order were matched to different options, that would serve as an evidence that the order of the nested epistemic expression affects how people interpret its meaning.

For the experimental item illustrated in Figure 3-16, the epistemic modals in question were "certainly" and "may", while other experimental items may have a different combination of epistemic modals. In this study, we selected six pairs of nested modals, which were "definitely and may", "definitely and might", "certainly and might", "certainly and may", "probably and might", and "probably and may". Each modal combination contained two epistemic modals, one expressing higher probability than the other, and each modal combination appeared in eight experimental items.

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

Prompt: A group of 10 friends are at Disneyland for the day. They're trying to decide which ride to go on first. After much deliberation, it's impossible that they'll go on Thunder Mountain first.

Option 1: 5 friends do want to ride Thunder Mountain first.

Option 2: 9 friends do not want to ride Thunder Mountain first.

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

3.6.3 Procedure

Participants used personal computers to access the web page hosting this online study, which started with demographic questions followed by the instruction and two practice trials. The practice trails mimicked the format of the experimental items in which there was a prompt followed by two options for participants to choose from. The prompt and the options were on the same page, and participants were instructed to select one of the two options that was most consistent with the information in the prompt. After the practice trials, participants hit a button to proceed. They were randomly assigned to one list of experimental items, and the first experimental item in the list was presented on the computer monitor screen. Participants read the prompt and selected the option the same way as they did in the practice trails. After that, they hit the proceed button to reveal the next experimental item. Once the new experimental item

revealed itself, there was no way to return to the previous item. The study ended after the participant completed the task for all 96 the items in the list.

3.6.4 Data analysis

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

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

Choosing the first option = 1 + condition + (1 + condition|subject)

+(1 + condition|item)

family = bernoulli (link = ”logit”)

We built two models of the above structure for items in the single-modal conditions and double-modal conditions separately. The fixed effects were dummy-coded, with the High and High-Low condition being the baseline for the single-modal conditions and nested-modal conditions respectively. The data and script used for statistical analysis of this study is available in GitHub[12].

3.6.5 Results

The frequency of each option selected by participants in different experimental conditions is shown in Figure 3-18. For experimental items that only contained a single epistemic modals, when participants were shown a prompt in the High condition, they selected the option corresponding to the semantics of the High modal in 1083 trials, while for the other 69 trials, they selected the option corresponding to the semantics of the Low modal. When participants were shown a prompt in the Low condition, they selected the option corresponding to the High modal in 587 of the trials, and selected the option corresponding to the Low modal in 565 of the trials.

[12] https://github.com/PON2020/Nested_Epistemic_Expressions_Further_Inquiry/tree/main/Data_Analysis

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

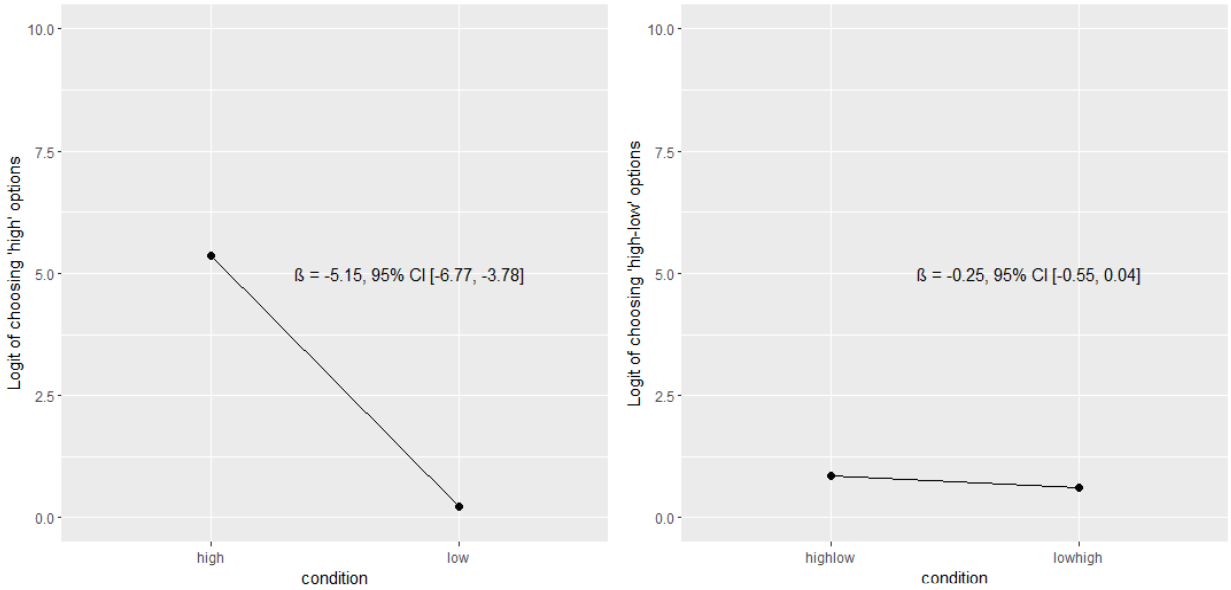
(a) Frequency of selecting each option in the single-modal conditions

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

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

Figure 3-18 Experiment 6: Frequency of the options selected in different experimental conditions

The difference between High and Low condition in the probability of choosing the option corresponding to the High modal is further illustrated in Figure 3-19(a) and Table 3.6-1. When reading the prompt that contained a single epistemic modal expressing lower probability, participants were less likely to select the option that was semantically equivalent to the High modal prompt (β = -5.06, 95% CI = [-6.66, -3.72]) compared with the baseline condition in which the participants were shown a prompt that contained a single epistemic modal expressing higher probability. The difference between High and Low condition in the probability of choosing the "High" option is further illustrated in Figure 3-19(a) as a noticeable negative slope.



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

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

Figure 3-19 Experiment 6: Fitting the logistic regression models for the single and nested conditions

For the experimental items that contained nested epistemic expressions, when participants were shown a prompt that contained two nested modals in the High-Low order, they selected the option corresponding to the semantics of the High-Low modal combination in 766 trials, while for the other 386 trials, they selected the option corresponding to the semantics of the Low-High combination. When participants were shown a prompt that contained two nested modals in the Low-High order, they selected the option corresponding to the High-Low combination in 709 trials, and selected the option corresponding to the Low-High combination in 443 trials. The output of the logistic regression model showed that the probability of selecting the option semantically equivalent to a nested expression in High-Low word order remained nearly the same regardless of whether the participants was given a prompt containing a High-Low modal combination or Low-High modal combination (see Figure 3-19(b)). When participants were shown a prompt containing a nested epistemic expression in Low-High word order, they

were slightly less likely to select the option that was semantically equivalent to the High-Low modal combination, compared to the baseline condition in which a High-Low prompt was given. However, such a small difference between the High-Low condition and Low-High condition was not statistically meaningful, given that zero was included in the credibility interval of the slope (β = -0.27, 95% CI = [-0.58, 0.03], see Table 3.6-2).

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	5.32	0.68	4.15	6.80
Low	-5.06	0.75	-6.66	-3.72

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

Table 3.6-1 Experiment 6: Estimate of intercept and slopes for single-modal data with the High condition as the baseline

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.88	0.15	0.58	1.19
Low-High	-0.27	0.16	-0.58	0.03

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

Table 3.6-2 Experiment 6: Estimate of intercept and slopes for nested-modal data with the High-Low condition as the baseline

3.6.6 Discussion

In this experiment, we investigated the processing of nested epistemic expressions by creating different interpretations of the expression and asking participants to select the interpretation that best represents their own understanding of the expression. Among the options, there was one interpretation that was supposed to be the correct interpretation based on the scope

account. If participants constantly selected the theoretically preferred interpretation, and also if the difference in the word order led to a change in the preferred option, we would be confident that the underlying mechanism of processing nested epistemic expression is in line with the scope account.

According to Moss (2015), the semantics of an epistemic expression can be modeled as having mental committee members vote for the acceptance of a statement, for example, "certainly S" means "every committee member accepts S", while "may S" means that "some committee members accept S". We know from Experiment 4 and Experiment 5 that "certainly" did not express a hundred percent probability, so the semantics of "certainly S" should be better interpreted as "the vast majority of the members in the committee accepts S". We further quantified the numeric representation of each epistemic modal under the mental committee framework (see Figure 3-17) and examined the extent to which participants' interpretation of the epistemic expression was in line with the mental committee representation.

Based on the mental committee framework, for statements containing only one epistemic modal, the strength of that modal is reflected from the number of the committee members who vote for a certain proposal. For example, "Mrs. Gherkin's blueberry pie is certainly going to win the purple ribbon" means "8 ladies on the church committee are going to vote for Mrs. Gherkin's blueberry pie", while " Mrs. Gherkin's blueberry pie may win the purple ribbon" means "6 ladies on the church committee are going to vote for Mrs. Gherkin's blueberry pie". The probability expressed by the modal "certainly" is higher than the probability expressed by the modal "may", and such a difference is represented by more votes in the scenario corresponding to the meaning of "certainly". The result of this experiment showed that participants were more likely to match the scenario "8 ladies on the church committee are going to vote for Mrs. Gherkin's blueberry

pie" to the statement "Mrs. Gherkin's blueberry pie is certainly going to win the purple ribbon" than to the statement "Mrs. Gherkin's blueberry pie may win the purple ribbon". This suggested that the semantic difference between the modal "certainly" and "may" can be represented using the mental committee framework, and such a representation was in line with how participants processed the meaning of the modal "certainly" and "may". The next question is whether or not participants' processing of nested epistemic expressions is also consistent with the prediction of the mental committee framework.

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

Similar to the previous experiments, the findings in this experiment is more in line with the prediction of the good-enough processing framework (Christianson, Williams, Zacks, &

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

References

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

Battistella, E. L. (2013). The syntax of the double modal construction. *LINGUISTICA atlantica*, 17, 19–44.

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

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

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

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

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.

De Haan, F. (2001). The Place of Inference within the Evidential System. *International Journal of American Linguistics*, 67(2), 193–219.

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

Elsman, M., & Dubinsky, S. (2009). Double modal syntactic patterns as single modal interactions. *University of Pennsylvania Working Papers in Linguistics*, 15(1), 10.

Epley, N., & Gilovich, T. (2006). The anchoring-and-adjustment heuristic: Why the adjustments are insufficient. *Psychological science*, 17(4), 311–318.

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

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

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

Gisborne, N., & Holmes, J. (2007). A history of English evidential verbs of appearance. *English Language and Linguistics*, 11(1), 1–29.

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

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

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, 5–1.

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

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

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas—part ii. *Language and Linguistics Compass*, 10(11), 591–613.

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

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

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.

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.

Appendix

This appendix provides more examples of the experimental items used in each experiment. A full list of experimental items for each experiment is available in the GitHub repository, the link to which can be found in the footnotes of this chapter.

Examples of the Stimuli in Experiment 1			
	Dialogues	Questions	Conditions
1	"Do you know where Janet is? The boss is looking for her." "She must probably be using the restroom."	How likely is it that Janet is using the restroom?	high low
2	"The train is stopping. Where are we now?" "We probably must be around Richmond."	How likely is it that the speakers are around Richmond?	low high
3	"Whose car is it? I haven't seen it before." "It must be Professor Murray's."	How likely is it that the car is Professor Murray's?	high
4	"How much is Claudia's new purse?" "It is probably over a hundred dollars."	How likely is it that Claudia's new purse is over a hundred dollars?	low
5	"Where is my green cap?" "It would possibly be in the basement."	How likely is it that the cap is in the basement?	high low
6	"Whose hat is it?" "It possibly would be Miss Swift's hat."	How likely is it that the hat is Miss Swift's?	low high
7	"Where is my red pen?" "It would be in the yellow pencil box."	How likely is it that the red pen is in the yellow pencil box?	high
8	"What is the weather like in Taipei now?" "It is possibly windy."	How likely is it that the weather in Taipei now is windy?	low

Examples of the Stimuli in Experiment 2			
	Dialogues	Questions	Conditions
1	"What will mom buy from the supermarket? Some sea food?" "She definitely may buy some seafood."	How likely is it that mom will buy some seafood from the supermarket?	high low
2	"Tom is not happy these days. Does that have something to do with the election?" "It may definitely be the election that made him unhappy."	How likely is it that Tom is not happy because of the election?	low high
3	"This soup tastes weird. Is it star anise that causes such a taste?" "The soup definitely have some star anises in it."	How likely is it that the soup has some star anises in it?	high
4	"The apple pie tastes very nice. Where did Jane get the recipe? From her mom?" "Jane may have gotten the recipe from her mom."	How likely is it that Jane got the apple pie recipe from her mom?	low
5	"Anna's new hairstyle looks very pretty. Where did she get her hair cut? From the new salon downtown?" "She definitely might have gotten her hair cut from the new salon downtown."	How likely is it that Anna got her new hairstyle from the new salon downtown?	high low
6	"Jackson didn't come to school this morning. What did he do last night? Drinking?" "He might definitely have drunk alcohol last night."	How likely is it that Jackson drank some alcohol last night?	low high
7	"Ellen has moved to a new place. Is it because of the problem with her neighbor?" "It is definitely the case."	How likely is it that Ellen moved to a new place due to the problem she has with her neighbor?	high
8	"Bob hasn't arrived yet. I am wondering if there is heavy traffic on his way home." "There might be a traffic jam on Bob's way home."	How likely is it that there is a traffic jam on Bob's way home?	low

Examples of the Stimuli in Experiment 3			
	Dialogues	Questions	Conditions
1	"Do you know where Janet is? The boss is looking for her." "She must, the manager said, probably be using the restroom."	How likely is it that Janet is using the restroom?	high low
2	"The train is stopping. Where are we now?" "We probably, according to the announcement, must be around Richmond."	How likely is it that the speakers are around Richmond?	low high
3	"Whose car is it? I haven't seen it before." "Based on what I heard, it must be Professor Murray's."	How likely is it that the car is Professor Murray's?	high
4	"How much is Claudia's new purse?" "I guess it is probably over a hundred dollars."	How likely is it that Claudia's new purse costs over a hundred dollars?	low
5	"Where is my green cap?" "It would, the new janitor suggested, possibly be in the basement."	How likely is it that the cap is in the basement?	high low
6	"Whose hat is it?" "It possibly, the cleaner said, would be Miss Swift's hat."	How likely is it that the hat is Miss Swift's?	low high
7	"Where is my red pen?" "After we reorganized the office yesterday, it would be in the yellow pencil box."	How likely is it that the red pen is in the yellow pencil box?	high
8	"What is the weather like in Taipei now?" "According to the weather forecast, it is possibly windy."	How likely is it that the weather in Taipei now is windy?	low

Examples of the Stimuli in Experiment 4 and 5			
	Dialogues	Questions	Conditions
1	"Do you know where Janet is? The boss is looking for her." "Janet must probably be using the restroom."	How likely is it that Janet is using the restroom?	high low without parenthetical
2	"Tom is not happy these days. Does that have something to do with the election?" "The election, as his partner mentioned, may have made him unhappy."	How likely is it that Tom is not happy because of the election?	low with parenthetical
3	"Whose car is this? I haven't seen it before." "The car, based on what I heard, must be Professor Murray's."	How likely is it that the car is Professor Murray's?	high with parenthetical
4	"What does Henry do for work now?" "Henry might, I have heard, probably work at a coffee shop."	How likely is it that Henry works in a coffee shop?	low high with parenthetical
5	"Where is my green cap?" "Your cap would, the new janitor suggested, possibly be in the basement."	How likely is it that the cap is in the basement?	high low with parenthetical
6	"Whose hat is this?" "That hat is possibly Miss Swift's."	How likely is it that the hat is Miss Swift's?	low without parenthetical
7	"Where is my red pen?" "Your pen would be in the yellow pencil box."	How likely is it that the red pen is in the yellow pencil box?	high without parenthetical
8	"What will the weather be like if you travel to Taipei tomorrow?" "The weather possibly would be windy."	How likely is it that the weather in Taipei tomorrow will be windy?	low high without parenthetical

Examples of the Stimuli in Experiment 6			
	Prompts	Options	Conditions
1	10 club members vote on who to elect treasurer. At the end, they think that Flora is definitely going to become treasurer.	6 members think Flora is going to become treasurer. 8 members think Flora is going to become treasurer.	high
2	10 colleagues are trying to figure out where to have their bi-annual company dinner, after a discussion, it seems like they might go to Chili's.	6 members support going to Chili's. 8 members support going to Chili's.	low
3	The board is hiring a new secretary. 10 members meet to discuss a particular candidate, and in the end, it seems that Suzan certainly might get the job.	6 members claim they are certainly in favor of Suzan getting the job. 8 members claim they might be in favor of Suzan getting the job.	high low
4	10 men on the Elk Lodge committee are voting on the chili contest. After discussing the nuanced flavors of the various chilies, it seems that the jerky chili might probably win.	6 men agree that the jerky chili probably tastes the best. 7 men agree that the jerky chili might taste the best.	low high
5	The 10 Girl Scouts of Troop 47 are voting on the Girl of the Year. After discussing the merits of each girl's actions, it seems that Daisy is certainly going to receive the honor.	6 girls want Daisy to receive the honor. 8 girls want Daisy to receive the honor.	high
6	The 10 Boy Scouts of Troop 19 are voting on the best campground. After discussing the merits of each option, it seems that Elk Ridge Lake may be voted the best.	6 boys believe Elk Ridge Lake is the best. 8 boys believe Elk Ridge Lake is the best.	low
7	A group of 10 friends are deciding whether or not to buy tickets to the new Jonas Brothers concert. After further discussion, they certainly might go to the concert.	6 friends say they certainly support going to the concert. 8 friends say they might support going to the concert.	high low
8	A family of 10 is deciding where to go on vacation. They ultimately decide that they might definitely visit Hawaii.	6 family members say they are definitely going to vote for Hawaii. 8 family members say they might vote for Hawaii.	low high

4 General discussion

In the above six experiments, we examined the processing of nested epistemic modality focusing on how interlocutors interpret modally non-harmonic 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 the scope account tested in this study, the meaning of the second modal should be interpreted within the scope of the first modal (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). In the first five experiments, we focused specifically on the overall 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 was 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?

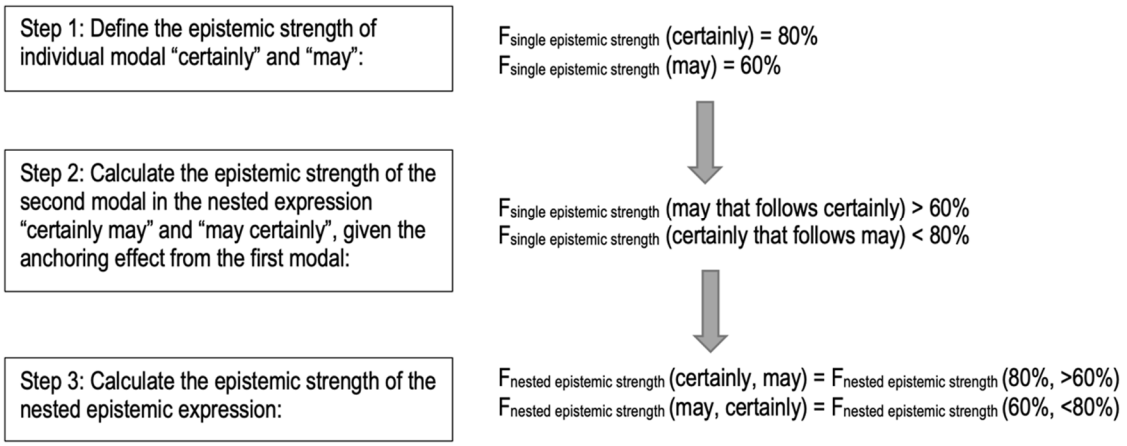


Figure 4-1 The computation of epistemic strength based on the scope account

The explanation we offered here treated the strength of the nested epistemic expression as a function that applies to the strength of each component modal. 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 strength of the expression as the output. As illustrated in Figure 4-2, suppose the strengths of individual modal "certainly" and "may" are 80 and 60 out of a hundred 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 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. It is important to note that the computational mechanism illustrated above is one of many possible ways in which the scope difference of the component modals maps onto the perceived probability of the embedded statement. In this study, we only tested this scope account due to its explicitness and its specific directional implications. 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 without fully pinning down the scope of the two modals or computing the meaning of the second modal under the influence of the first modal. The findings of the 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 non-harmonic (Lyons, 1977, p. 807). What we found was that the order of the two modals did not affect the probability rating score of the statement. As we can see from Section 3.1.5, the difference between the two nested modal conditions in probability rating was about two out of a hundred, 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 modal expressed very low probability ("may" and "might"). Results from the first two experiments suggested that when processing a statement embedded within two epistemic modals, one expressing high probability and the other expressing low probability, the processor 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 (Di Paolo, 1989), 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 an individual modal. For the previous two experiments in which the parenthetical elements were not included, significant differences were found between the high modal condition, 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.

Experiment four and five further examined the effect of word order and parenthetical element by including both factors in a two by two design, while the basic paradigm remained unchanged. The findings of the first three experiments were successfully replicated in Experiment 4 and 5. First of all, the strength of English epistemic modals formed a clear scale

from expressing high probability to intermediate probability and then to low probability. On the high end of the scale were words such as “certainly” and “definitely” expressing 80 to 90% probability, while on the low end of the scale, words such as “may” and “might” expressed 50 to 60% probability. The word “probably” expressed intermediate degree of probability lying in between the strength of “may” and “certainly”. The two modals in a nested epistemic expression were selected from different positions on the scale, and thus, if the scope of the two modals were processed, interlocutors would form different interpretations for the expressions with the same modal combination but in the opposite word order. What we found in experiment four and five was that there was no difference in the probability rating of the nested expressions in different word order. For example, the probability of the statement S in “certainly may S” and “may certainly S” had the same rating score. The lack of ordering effect held true regardless of whether the two modals were adjacent or separated by a parenthetical element. Moreover, the interaction effect between the presence of a parenthetical element and the epistemic strength of an individual modal has been confirmed. We found that adding a parenthetical element boosted the probability rating of the statement embedded within the scope of a single Low modal to a statistically significant extent. Interestingly, the effect of adding a parenthetical element was quite the opposite for the statements embedded within the scope of High modals- the presence of a parenthetical element slightly decreased the rating of the embedded statement.

The parenthetical elements used in Experiment 3, 4 and 5 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: 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, 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 based on the word order, such difference should be captured by this research paradigm when we averaged the rating scores across all the items. Instead of finding an order effect among the two nested-modal conditions, what we found in experiment 4 and 5 was that statements embedded within a single modal of low epistemic strength were rated higher when a parenthetical evidential was present, while statements embedded within a single modal of high epistemic strength were rated lower when a parenthetical evidential was present. 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 findings of our experiments suggested that comprehenders would assign higher

probability to “the soup has star anise” if they heard the first sentence, rather than the second sentence. On the other hand, comprehenders would assign higher probability to “the soup has star anise” if they heard “The soup must have some star anise in it” compared with “Based on my experience, the soup must have some star anise in it”.

Linguistic theories have catalogued 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”, but not so when they hear “according to a child who always lies, the soup may have some star anise in it”. Different information sources are associated with different epistemic strengths, 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 utterence “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 evidential as expressing a median degree of credibility. Given that the epistemic modal “may” in this sentence expresses low probability, the inclusion of the parenthetical evidential boosts the overall probability of the

statement. It follows that for an epistemic expression containing a single modal expressing very high probability like “certainly” and “must”, adding the parenthetical evidential decreases the overall strength of the expression.

The last experiment investigated the processing of nested epistemic expressions by using a different paradigm. Instead of asking participants to rate the probability of a statement based on their understanding of a nested epistemic expression, the new paradigm provided participants with two interpretations of a nested expression, and let participants choose the option that was closer to their own interpretation. Among these two options, one interpretation was supposed to be the correct interpretation based on a scope account, which modeled the meaning of epistemic modals as voting scenarios among an imagined committee (Moss, 2015). Based on this mental committee framework, “certainly S” means “eight out of ten committee members accept S”, while “may S” means “six out of ten committee members accept S” (“S” here stands for a statement). We found that participants’ interpretation of the single epistemic modal was largely in line with the prediction of the mental committee framework, for example, participants were more likely to match the scenario “eight out of ten committee members accept S” to a prompt containing “certainly S” than a prompt containing “may S”.

Interestingly, for expressions like “may S”, participants in half of the total trials interpreted it as saying “eight out of ten committee members accept S”, which was supposed to be the interpretation for “certainly S”. This reflected the entailment relation among items on the scale of epistemic strength. The modal “certainly” and “may” are on the same scale of epistemic strength with “certainly” being on the high end and “may” on the low end. Items on the same scale are not only qualitatively similar (Gazdar, 1979) and quantitatively comparable; they also carry specific semantic and pragmatic implications, as Levinson (1983, p. 134) nicely

summarized “the semantic content of lower items on a scale is compatible with the truth of higher items obtaining, and the inference that higher items do not in fact obtain is merely an implicature”. Semantically speaking, an utterance containing an element higher on the scale entails an utterance containing an element lower on the scale (Gazdar, 1979; Horn, 1972; Van der Auwera, 1996). For example, “certain” is higher on the epistemic scale than “possible”, and the sentence “it is certain that he is late” logically entails that “it is possible that he is late”, but not the other way round. This explained why when shown the prompt containing “may S”, participants interpreted it as saying “eight out of ten committee members accept S” in about half of the trails. This is because, if eight out of ten committee member accept S, it is certainly the case that S, which entails “may S”. On the other hand, the option “six out of ten committee members accept S” means “it may be the case that S”, which does not semantically entail “certainly S”. That is why for the prompt containing “certainly S”, only a small number (about 6%) of trails interpreted it as “six out of ten committee members accept S”.

Though semantically speaking, the proposition “certainly S” and “may S” can be true at the same time, pragmatically speaking, saying “may S” implies the negation of “certainly S” (Gazdar, 1979; Verstraete, 2005). The force choice task in this experiment seemed to encourage participants to form more semantically driven interpretations when processing epistemic expressions, however, as long as participants’ preferred choice differed across experimental conditions, that would serve as an evidence that participants processed epistemic expressions differently in different experimental conditions. We found that for epistemic expressions containing one epistemic modal, participants interpreted their meaning differently depending on whether the modal had a higher epistemic strength or lower epistemic strength. However, for

nested epistemic expressions, regardless of the word order of the embedded modals, no difference was observed in participants' interpretation.

For a nested expression like "certainly may S", in the majority of the trails (about 66%) participants interpreted it as saying "eight out of ten committee members accept that it may be S", and the minority (about 34%) interpreted it as saying "six out of ten committee members accept that it is certainly S". The same pattern was observed for the nested expression in the opposite word order. For expressions like "may certainly S" the majority of the trails also favored the interpretation saying "eight out of ten committee members accept that it may be S". The logistic regression model revealed that the probability of choosing the interpretation corresponding to the High-Low modal pair remained the same regardless of whether the participants were shown a prompt containing a High-Low expression or a prompt containing a Low-High expression. This suggested that the order of the embedded modals did not affect the semantic interpretation of the nested epistemic expressions.

The focus of this research project is to investigate whether or not the processing of nested epistemic expressions follows the prescription of formal linguistic theories, according to which one component modal in a nested expression should be interpreted within the scope of the other modal (Lyons, 1977; Moss, 2015). 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 we tested in this project 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 do not have the same meaning as the modal operators differ in their priority,

which is reflected from the relative position of the modal adverb and modal auxiliary in the syntactic hierarchy (Potsdam, 1998). 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, 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 project 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.

Since the focus of this study is on the processing mechanism of a specific type of modal constructions, it does not directly inform the theoretical representations of nested modals in syntactic and semantic architectures. The competence of analyzing the scopes of modal operators should be well preserved among interlocutors, and what we found is that in a casual communicative environment, interlocutors tend to process nested epistemic expressions in a good-enough manner, rather than strictly follow the linguistic representations that are assumed to underlie the forms. This research echoes the good-enough processing framework that has been proposed in human sentence processing (Christianson et al., 2006; Ferreira et al., 2002, 2001; Ferreira & Lowder, 2016). It is interesting to think about the extent to which the good-enough processing underlies the comprehension of nested epistemic expressions, especially in cases where the scope of the modals are more explicit. For example, future research could compare how people interpret the syntactically elaborated version of the nested expressions such as "it is certainly the case that he may have forgotten" and "it may be the case that he certainly has forgotten". In this case, the surface form of the sentence calls attention to the different scopes of the modals, and if a lack of order effect is still observed, we would conclude that people generally treat instances of non-harmonic combinations of epistemic modals as conveying moderate uncertainty across the board. Alternatively, if interlocutors' interpretations of the elaborated versions show sensitivity to scope, this may suggest some idiosyncratic features of the nested expressions like "may certainly" that prompt the shallow processing of the scope.

Future research using online measures to investigate the processing of nested epistemic expressions would also be an important complement to this study. For all the six experiments reported in the previous chapter, we measured participants' interpretation of the nested expressions as the output of their processing of either visually or acoustically presented language

input. It is not clear how the nested constructions are incrementally processed, specifically, the processing load associated with the epistemic modals. For example, given that nested epistemic expressions are much less frequent compared with the use of a single epistemic modal, does the second epistemic modal require more processing effort than the first modal? If the second modal is associated with an increased processing load, does the word category of that modal, being either an auxiliary or an adverb, further modulate the ease of processing? Moreover, how does the frequency of the modal combination influence the processing of the two modals? Online measures such as eyetracking and recording of Event-Related Potentials (ERPs) could be employed to explore some of the above questions in future work.

Last, but not the least, it is interesting to further investigate the cross-linguistic generalization of nested epistemic expressions, focusing on their structures, frequency of usage and the way they are processed in casual conversation. This dissertation initiates the research on processing nested epistemic expressions with English being the language under investigation. It is likely that the nested structure of two epistemic modals in a single clause can be observed cross-linguistically with similar or different syntactic configurations. Studying the processing of nested epistemic expressions in other languages sheds light on our growing knowledge of linguistic diversity and unity.

References

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.

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

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

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

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

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

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

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

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

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. Academic Press. Retrieved from https://books.google.com/books?id=CtnWAAAAMAAJ

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

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

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

Horn, L. R. (1972). *On the semantic properties of logical operators in English*. University of California, Los Angeles.

Levinson, S. (1983). *Pragmatics.* 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, 5–1.

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.

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

Putnam, H. (1975). The meaning of 'meaning'. In *Philosophical Papers* (Vol. 2), Cambridge: Cambridge University Press. 215–271

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

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

Rooryck, J. (2001). State of the article: Evidentiality part i. *GLOT international*, 5(4), 125–133.

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

Van der Auwera, J. (1996). Modality: the three-layered scalar square. *Journal of semantics*, 13(3), 181–195.

Verstraete, J. C. (2005). Scalar quantity implicatures and the interpretation of modality: Problems in the deontic domain. *Journal of pragmatics*, 37(9), 1401–1418.

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.

5 Conclusion

The findings of this study provide answers to the research questions listed in Section 2.6. In terms of native English speaker's knowledge about the strength of various epistemic modals, this research in general confirmed the proposed three-point scale (Horn, 1972; Halliday,1970; Holmes,1982; Halliday and Matthiessen, 2004), with only a few exceptions. For example, while Halliday and Matthiessen (2004) claimed that the modal "would" expressed intermediate degree of probability, we found that English speakers nowadays interpret the word "would" as expressing high probability, similar to the epistemic modal "must". Our findings also suggested that the epistemic strength of a modal can be represented numerically. Modals on the high end of the scale, such as "certainly" and "definitely" express over 80% probability, while modals on the low end of the scale, such as "may" and "might" express around 60% probability. The modal "probably" consistently lies in the middle of the scale, expressing about 70% probability.

In terms of the compositionality of the nested epistemic expression, when the component modals differ in their epistemic strength, the overall strength of the nested expression lies in between the epistemic strength of the component modals. When processing non-harmonic expressions, the processor needs to take into consideration the epistemic strength of each component modal. To be more specific, the strength of the first modal lingers in memory, and is mixed with the epistemic strength of the second modal when it is encountered. The parser then normalizes the mixed epistemic strength as an indicator of general uncertainty without further pinning down the scope of each modal. This mechanism underlies the processing of nested epistemic expressions regardless of whether the component modals are adjacent or separated by other words in between.

The findings of this research challenge the scope account of nested epistemic expressions in language processing (Lyons, 1977, p. 808; Moss, 2015; Potsdam, 1998), and suggest a holistic processing mechanism in line with the "good enough" processing framework (Christianson, Williams, Zacks, & Ferreira, 2006; Ferreira, Bailey, & Ferraro, 2002; Ferreira, Christianson, & Hollingworth, 2001; Ferreira & Lowder, 2016). This study sets the stage for the use of measures such as eyetracking and recording of Event-Related Potentials (ERPs) to examine nested epistemic expressions in online processing and interpretation.

Reference

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.

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

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

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

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

Halliday, M.A.K. (1970). Functional Diversity in Language as Seen from a Consideration of Modality and Mood in English. *Foundations of Language* 6: 322-61.

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

Horn, L. R. (1972). *On the semantic properties of logical operators in English*. University of California, Los Angeles.

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

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

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