Talking about higher-order uncertainty

Daniel Lassiter, Stanford Linguistics


1 Introduction

Theories of the lexical semantics of modal language live in the shadow of modal logic. Being a junior partner in this relationship has many advantages for us linguists: we get to make use of technical results that come from modal logic, and we benefit from the linguistic insights of smart philosophers, many of whom might not bother with natural language if they didn’t think that modal expressions hold some philosophical interest. Still, the technical apparatus of modal logic was developed by people whose main academic interest was definitely not the lexical semantics of one corner of the vocabulary of a single Germanic dialect. For those of us who do (for whatever reason) care deeply about the vocabulary of English for its own sake, we should of course be cautious about borrowing in toto theories that were designed to elucidate philosophical issues—issues whose import does not depend on the quirks of any natural language. For each such case, it is important to examine the borrowed theory carefully in order to ensure that its empirical predictions make sense of the grammatical and inferential properties of the item in question.

Lauri Karttunen was perhaps the first linguist to make this point clearly and forcefully, in his “Possible and Must” (Karttunen 1972). He did so by exploring an extension of Hintikka’s (1962) influential logic of knowledge and belief to English expressions of epistemic modality—possible, may, perhaps, must, and have to among others—and showing that the empirical predictions of the extended theory failed to match up with the intuitive interpretations of the English expressions in various ways. Lauri was careful to emphasize that these findings only impugned the application of standard systems of modal logic to the lexical semantics of epistemic modals in English: “I will argue that the principles on which our use of possible and must in ordinary language appears to be based are different from any of the well known standard systems that have been studied by logicians” (p.3). The same systems could still be relevant to the lexical semantics of other items, and to the elucidation of philosophical issues that are independent of the semantics of English: “[T]his is an interesting fact about ordinary language, but it should not be misconstrued as evidence against other modal systems” (ibid.).

Lauri’s insights in “Possible and Must” have generated a significant literature, especially around his novel observation that the intuitive interpretation of must is weaker than that of the necessity operator in standard modal logics of knowledge. A few of the papers in this still-growing collection are Groenendijk & Stokhof 1975; Kratzer 1991; von Fintel & Gillies 2010; Lassiter 2016; Mandelkern 2016; Goodhue 2018. Thanks to “Possible and Must”, there is now a consensus among linguists that must differs in some important ways from the □ operator of modal logic S5—though
there is still no consensus about exactly what the difference is. S5 is the most obvious way to interpret linguists’ usual gloss of *must, might, perhaps, possible* etc. as ∀- or ∃-quantifiers over a “set of epistemically possible worlds” which contains the actual world. Under certain common (though not uncontroversial) assumptions, these quantifiers correspond to the □ or ◇ of a normal modal logic where accessibility is an equivalence relation (reflexive, transitive, and symmetric).

In this paper I want to talk about another property of epistemic modals in English that apparently differentiates them from the S5 modal operators: the fact that they can be nested in a non-trivial way. A key property of S5 is that iterated modals reduce to the innermost. For example,

\[ \square \square \phi \Leftrightarrow \square \phi \]
\[ \square \diamond \phi \Leftrightarrow \diamond \phi \]
\[ \diamond \diamond \phi \Leftrightarrow \square \phi , \]

and so on. In S5, piling modal operators on top of an already modalized formula has no semantic effect. This is due to the fact that accessibility in S5 is an equivalence relation. The worlds accessible from \( w \) are exactly the same as the worlds accessible from any world accessible from \( w \). If all of the worlds you can access in one step from \( w \) verify \( \phi \), then all of the worlds you can access in two steps will also verify \( \phi \). So, \( \square \phi \) holds at \( w \) if and only if \( \square \square \phi \) does.

However, as Moss (2015) discusses in detail, this property does not seem to characterize English accurately. In examples like those in (1), there are two epistemic expressions, one of which takes scope immediately above the other. Yet both appear to make a separate contribution to the meaning of the sentence, in violation of the S5 principle that the modal that takes wider scope should be vacuous.

(1) a. This book is an odyssey, a journey up through the mists of time from the remote past. It .... explores what must perhaps be the most fundamental of all questions - who we are. (Dunbar 2004: 7)
   b. Alice is a likely hire, and Bob might be a likely hire. (Moss 2015)
   c. “At this point, it’s probably unlikely that we’ll trade Doc,” Ricciardi said Tuesday. (San Jose Mercury News, 7/21/2009)

This paper is an interim report on a project with two main components. Taking a probabilistic perspective on epistemic modality, I attempt to connect the interpretation of nested epistemics with theories of uncertainty that have been developed in philosophy and psychology, but without a focus on natural language. I then report on a study that explores the distribution of examples following the pattern in (1) in a large corpus of *New York Times* articles. I suggest that the examples of nested epistemics found in this corpus vary in the kind of uncertainty involved: in most of them, one or both of the modals is not epistemic (talking about information) but rather objective (‘worldly’, ‘stochastic’: talking about chance as a feature of the world itself). If this is right, then many or most naturalistic examples that follow the pattern in (1) do not technically constitute counter-examples to the S5 reduction property, because one or both of the modal operators is not being interpreted epistemically. While there are a few examples where true higher-order uncertainty may be implicated, I’ll suggest that the lower modal in these examples may be interpreted by ‘objectivizing’ the speaker’s own information, treating it as a feature of the world.
If the tentative generalizations floated here about the interpretation of nested epistemics hold up, then we have a situation that is interestingly different from the cases that Lauri discussed in his groundbreaking 1972 paper. There, Lauri showed that certain intuitions derived from the study of epistemic logic were misleading and should be discarded when doing modal semantics for English. Here, we have a case where what appears to be a straightforward counter-example to the S5 reduction property—cases where such modals are nested but both make a semantic contribution—may not constitute a counter-example after all. Instead, they point to an interesting ambiguity in the kind of uncertainty that “epistemic” expressions trade in.

2 A compositional puzzle

When dealing with examples like (1), these are some key questions that we need to resolve:

A) What do these sentences mean?
B) How can we define a compositional interpretation procedure for the meanings in (A)?
C) How do the meanings of uncertainty expressions relate to what we know about the psychological representation of uncertainty?

Compositionality is a crucial principle that should guide our investigation. That is, it is desirable to have a single theory of the meaning potential that probably contributes in examples (2a) and (2b)—not just methodologically, but also psychologically because such an analysis helps to explain the productive use and interpretation of modal language.

(2) a. The time is now near at hand, which must probably determine whether Americans are to be free men or slaves ... (G. Washington)

b. This time will probably determine whether Americans are to be free men or slaves.

I will assume that unembedded probable, probably, and likely are interpreted as indicating that the proposition that it embeds has high probability (Yalcin 2005, 2007, 2010; Swanson 2006, 2015; Lassiter 2010, 2015, 2017). Roughly:

(3) probably $\phi \leftrightarrow P(\phi) > .5$

A probabilistic interpretation is also appropriate for complex expressions built from probable and likely such as improbable, very likely, probability, and more likely than not: see Lassiter 2015, 2017 for compositional details.

In addition, I will assume—rather more controversially—that certain other epistemic expressions including possible, might, must and certain are also interpreted directly in terms of probabilities. Possible and might mean roughly probability greater than 0, and certain and must mean probability 1 or something close to it. For a detailed analysis of these expressions and their relationship to probable and likely see Lassiter 2015, 2016, 2017. If you do not like the latter assumption, feel free

---

1 Example borrowed from Moss (2015). The quote is from a speech delivered on August 26, 1776, before the Battle of Long Island.
to think of it as a simplification for expository purposes, holding place for a more elaborate theory which captures the probabilistic entailments of these expressions in a different way.

The question, then, is how to embed the interpretation of *probably* etc. in a way that correctly predicts the interpretation of Washington’s utterance. If we treat the ‘*P*’ as a parameter of interpretation (as in Yalcin 2007), the choice of probability measure is independent of the choice of evaluation world. If we want these sentences to interact non-trivially with embedding operators, then, we have to do something more complicated than the usual trick of abstracting over the evaluation world. Alternatively, we could look for a way to make the choice of probability measure sensitive to the choice of evaluation world. That is, if we had a theory according to which the meaning of Washington’s “probably” depends on features of the world—rather than the utterance context or some other global feature of conversation—then we could treat *probably* φ as denoting a proposition. *Must* in Washington’s statement would then take as its argument an ordinary proposition, the set of worlds

\[ \{ w \mid P_w([\phi]^{c.w}) > .5 \} .\]

Here \( P_w \) is the hypothesized world-dependent probability measure, and \( \phi \) is the argument of *probably* (*that this time determines whether Americans are to be free men or slaves*).

One reason for taking this kind of approach seriously is that there is a well-motivated theoretical distinction among multiple kinds of probability, only one of which is connected to information states. The other two that I will discuss are best defined in terms of objective features of the world. These distinctions have been developed theoretically and motivated empirically in philosophy and cognitive psychology, and they have a linguistic trace as well. If it turns out that epistemic modals nested under other epistemic modals are interpreted according to the world-dependent kind of probability, we have a simple solution to the compositional puzzle: what *must* embeds in (2a) is an ordinary proposition.

The rest of this paper will explore the idea that nested epistemics may be interpreted in this way. I’ll begin with a discussion of how the distinction has been made in philosophy and psychology, and then use this discussion to consider the classification of some examples drawn from a large corpus of news text.

### 3 Two kinds of probability

Philosophers of probability have distinguished between subjective and objective probability (e.g., Hacking 1975; Lewis 1980; Gillies 2000). Subjective probability, also known as “credence” or “degree of belief”, is a property of individuals’ psychological states. Pure subjective uncertainty might involve questions about unique events that have a determinate answer, but for which we lack certain key information. For example, I have no idea whether Plato was taller than Aristotle, but I do know that there is a determinate answer to this question. My uncertainty is of the purely subjective type.

Objective probability, also known as “chance” or “propensity”, is a non-psychological property of the world. A truly stochastic event—say, the time that it takes for a certain radioactive particle to decay—is intrinsically chancy, independent of what anyone believes about it, or indeed whether any agents exist. Along similar lines, the particle might be said to have a certain propensity to
decay within a given period of time. While the extension of this kind of reasoning to non-quantum events is metaphysically controversial, it is natural to think of certain kinds of macroscopic events as intrinsically random as well. A flip of a fair coin has a 0.5 chance of coming up heads, independent of what anyone thinks about it; this is a fact about the world. If I draw a ball at random from a well-mixed urn with 3 red and 7 green balls in it, there is a 0.3 chance of getting a red ball. The roll of a die, a deal of a card from a well-shuffled pack, and the result of a future sporting event whose outcome depends on many chance factors might all be thought to have a similar character. Naturally, we can also have mixed uncertainty: for instance, I might be uncertain about whether a given coin was fair, so that I am subjectively uncertain about the value of a certain objective chance (Lewis 1980).

One way to gain an intuitive feel for the distinction is to imagine yourself being in possession of all immediately causally relevant information about an event. For example, suppose that you were in immediate possession of all relevant information about the factors influencing an event’s outcome. The only thing that you do not have direct observational evidence of is the actual outcome. Would you still have any uncertainty about the outcome? If not, you are dealing with a case of purely subjective probability. Otherwise, the situation includes at least some degree of inherent stochasticity, i.e., objective chance. Of course, the problem with this heuristic is that it relies on the intuitive notion of “immediately causally relevant” information. If the event in question involves a coin flip or weather conditions, a layperson and a physicist might resolve this expression in radically different ways. But, for most of us, it is natural to suppose that there would be a certain amount of inherent stochasticity in coin flips, draws from an urn, the outcomes of sporting events, atmospheric events, and the like.

This latter point brings out one of the reasons why the extent of objective chance is controversial in philosophy—does it exist beyond the quantum level?—or indeed whether it is strictly needed at all. I want to sidestep such metaphysical questions by taking a psychological perspective on meaning: the question I am interested in is not whether coin flips really involve objective chance, but whether people believe that it does, and whether such a belief influences the way that they communicate about uncertainty. So, we should turn to the psychological literature for clues about whether and to what extent people are sensitive to such a distinction in reasoning, decision-making, and communication.

Psychologists have distinguished two kinds of uncertainty that mirror the philosophical discussion. “Singular external” (Kahneman & Tversky 1982) or “aleatory” (Ülkümen, Fox & Malle 2015) uncertainty involves events that are considered to be intrinsically random and essentially unpredictable. This type of uncertainty is intuitively the same as objective chance, as long as we are willing to extend the latter concept well beyond the quantum level. They argue that our knowledge of external uncertainties is generally due to reasoning about specific cases, e.g., about the physical characteristics of a certain coin, or about the likely trajectory of an arrow fired from a bow pointed in a particular direction, under specific atmospheric conditions. Corresponding to subjective probability is what psychologists have called “internal” (Kahneman & Tversky 1982) or “epistemic” (Ülkümen et al. 2015) uncertainty.² This is the kind of uncertainty that we have about

² The label “epistemic” is unfortunate, since it is really subjective beliefs that are relevant, regardless of the evidence on which they are based constitutes knowledge.
in-principle knowable features of the world, such as the length of the Nile, or Plato and Aristotle’s relative heights.

In any case, given the proliferation of terminology in this domain, it will be useful to declare the labels that I will use at this point.

- **Objective probability**, a world-bound property of events involving inherent randomness
  - abbreviation: “O-probability” or just “O”
  - cf.: chance, propensity, singular external uncertainty, aleatory uncertainty, stochasticity, randomness

- **Subjective probability**, relating to an agent’s information about an in-principle knowable event
  - abbreviation: “S-probability” or just “S”
  - cf.: credence, degree of belief, internal uncertainty, epistemic uncertainty

(The best labels would perhaps be “chance” and “credence”, which are widely used in the philosophical literature on probability, but these will be less useful as abbreviations later on since they begin with the same letter.)

Bayesian qualms notwithstanding, there is considerable evidence for the linguistic and psychological reality of this distinction. On the psychological side, we have, for example, decision-making experiments which suggest that the distinction influences choice behavior. People prefer to place bets on events whose objective chance is known, rather than on those which involve pure subjective uncertainty or a mixture of subjective uncertainty and objective chance. For instance, Ellsberg (1961) famously observed that people generally prefer to bet on a binary event whose objective probability is known to be 50% rather than an event about which they have no information about objective probabilities, so that it could be anywhere in the \([0, 1]\) range—even if they are indifferent between the two possible outcomes in both cases.

While this example involves a mixture of subjective and objective uncertainty, it is possible to separate the two cleanly by considering how people’s betting behavior is affected by whether the chancy event has already occurred. In general, when given the choice between placing a bet on an event with known chances before or after it has occurred, people prefer to bet in advance (Brun & Teigen 1990). This is surprising: we would expect a strict Bayesian to be indifferent, since the subjective probability of winning the bet remains the same if there is no prospect of gaining information about the outcome once it has occurred. This result suggests that there is a preference for betting under pure O-uncertainty rather than pure S-uncertainty, and so, *a fortiori*, that people make this distinction psychologically. Relatedly, Rothbart & Snyder (1970) found that experimental participants had greater confidence in their predictions about the roll of a die if the die had not yet been rolled. Their participants were also willing to bet more money on the outcome of the roll before it had happened.

The latter findings are consonant with Kahneman & Tversky’s (1982) claim that people believe in a sharp asymmetry between past and future: past events are in-principle knowable, but future events—unless they are impossible or inevitable—involves external uncertainty. Coming from a
different perspective, Lewis (1980) likewise makes a past/future asymmetry central to his theory of the relationship between subjective probability and objective chance. Note that the temporal dimension is also important in resolving what is “in-principle knowable”. As a rule, past events are in-principle knowable, and uncertainty about them is of the purely subjective type. Future events may or may not be chancy: for instance, there is an objective chance of rain tomorrow, but no chance that $3 + 3$ will fail to equal six tomorrow. However, many interesting future events are chancy, and involve a mixture of subjective and objective uncertainty.

Turning to linguistic traces of the subjective/objective distinction, Ülkümen et al. (2015) addressed this question with a corpus study of uncertainty expressions in *New York Times* articles published in 2008-2009. They extracted all instances of a predetermined list of phrases from these articles, including “sure”, “confident”, “certain”, “chance”, “likely”, “likelihood”, and “probability”, and asked annotators to rate them on a number of dimensions including perspective, control over the event’s outcome, source of uncertainty, and locus of uncertainty (internal or external). An interesting finding was that uncertainty expressions vary in whether they are typically used to discuss subjective or objective probabilities. The most interesting result was that there was a clear preference for *likely, chance, likelihood*, and *probability* when the locus of uncertainty was external (objective), and for *sure, confident, certain, chance* when the locus of uncertainty was internal (subjective). While the result was far from categorical, it was highly reliable.

This aspect of Ülkümen et al.’s (2015) study reveals two interesting facts. First, the subjective/objective distinction matters for communicating about uncertainty (and not only reasoning and decision-making). Second, the association between particular items and particular types of uncertainty is not categorical. While it is not clear how it should be explained semantically, it would be a mistake to hard-code this association into the semantics of the various expressions. For example, it is easy to find examples like (4) where *likely* expresses pure subjective uncertainty.

(4) a. An accident report into a light aircraft’s crash landing at East Midlands Airport earlier this year has revealed the most likely cause was an incorrectly fitted screw in the landing gear.

b. [T]he residential-scale reservoirs ... were likely used around 900 B.C. It’s likely that the systems were lined with a thick, clay “plaster” that allowed the areas to hold the water instead of seeping away.

Presumably the cause of the plane crash is a fixed fact about the world: investigators simply have subjective uncertainty about what this cause was, and an incorrectly fitted screw is the best hypothesis. Similarly, the way that certain reservoirs were constructed and used in 900 B.C. is clearly not a matter of chance: subjective uncertainty is all we can muster.

Contrariwise, it may also be that *certain* can express pure objective chance in examples such as “It is certain that everyone will eventually die’. (But not, of course, in the frame “$x$ is certain that ...”).) However, it is more difficult to show that this is strictly objective since someone who is in a position to assert the ubiquity of death will presumably also be in a state of maximal subjective confidence about the truth of the same proposition.
A third kind of probability

A third kind of probability that seems to be linguistically relevant is what I will call “distributional” probability, involving information about frequencies or proportions in a reference class. (Abbreviation: D-probability. This corresponds to what Kahneman & Tversky (1982) call “Distributional external” uncertainty.) At a high level, this type corresponds to some extent to frequentist probability, but I want to frame it in a slightly different way: D-probability is the kind that is relevant when we imagine ourselves constructing an ad hoc group of real-world individuals or events, and considering the distribution of properties within this group. D-probability can be operationalized as the proportion of individuals in the group that satisfies a certain predicate—or, equivalently, the objective chance that a randomly sampled member of the group would satisfy the predicate. The following web examples illustrate.

(5) a. The eight anti-seizure medications in the study came in 37 colors and four shapes. Overall, a change in color saw about a 20 percent additional likelihood that a prescription would lapse.

b. At the University of Nebraska College of Law [in 2006] ... [t]he probability that a Hispanic resident with the credentials of the average black admittee would be admitted was 43 percent.

D-probability readings seem to arise frequently when there is an indefinite within the scope of a probability expression. The indefinite is apparently not interpreted as an existential quantifier. For example, neither scopal order in (6) captures the interpretation of (5a), on either an objective or a subjective interpretation of the probability measure \( \text{prob} \). Letting \( b \) be the implied baseline from which “additional” deviation is being measured in this example:

(6) a. \( \exists x [\text{prescription}(x) \land \text{prob}(\text{lapse}(x)) = b + 0.20] \)

b. \( \text{prob}(\exists x [\text{prescription}(x) \land \text{lapse}(x)]) = b + 0.20 \)

The first reading in (6) deals with uncertainties involving some particular prescription; the second is about the probability that any prescription would lapse. Neither corresponds to the natural interpretation of (5a). Rather, this example is about the probability that a randomly selected member of a certain real-world class (those prescriptions, among those on which the study gathered data, for which the color of the pills had changed) had a certain property (not being renewed by the patient). Similarly, the claim being made in (5b) is that, among the class of Hispanic applicants who had the overall credentials of an average black admittee, 43% were admitted. This is obviously different from the claim that, for some Hispanic applicant \( x \) with certain credentials, the probability that \( x \) would be admitted was or is 43% (subjectively or objectively). It is also different from a claim about the probability that a Hispanic applicant (any Hispanic applicant) with these credentials would be admitted. The chance that there would be at least one such applicant admitted was presumably much higher prospectively, and retrospectively it is either 1 or 0. (This holds both objectively, and subjectively given that all relevant admissions data were available.)

It would be interesting to develop a compositional interpretation for distributional readings of probability expressions—building, perhaps, on theories that treat indefinites as denoting free variables or choice functions. I will not try to do this here, though. The crucial facts for current
purposes are the following. First, probability expressions in English can make reference to a third type of uncertainty which has not, to my knowledge, been studied previously from a linguistic perspective. Second, D-probability is, like O-probability, a non-psychological feature of the world—as real as rain, sleet, or snow once the relevant class of individuals has been specified. Since objective facts about the world are sufficient to determine the proportion of individuals of individuals in a certain class that satisfy a certain property, it is natural to expect the interpretation of distributional probability expressions to be indexed to the world of evaluation.

5 An informal corpus study of nested epistemics

Our main goal here is to understand the interpretation of examples involving nested epistemics, where one epistemic expression occurs in the immediate scope of another. I pointed out that a compositional interpretation of these examples would be very simple if we had a way to “objectivize” probabilities, tying them to the world of evaluation rather than some parameter of evaluation or other non-world-bound feature of interpretation. Now we have two such ways—objective probability and proportional probability—and some evidence that at least some “epistemic” expressions are genuinely ambiguous in whether they express subjective, objective, or proportional probability.

If we wish to explain embedded occurrences of epistemics in the most conservative possible way—with epistemic modals taking a propositional argument—this would suggest the following strong hypothesis: Embedded epistemics always have a world-bound interpretation. That is, we should find examples of modals indicating O-, D-, and S-uncertainty in matrix position, but those embedded under another epistemic expression should always have an O- or D-interpretation.

However, we do not expect singly embedded examples to show the full 3x3 typology of ambiguities that is in principle available. The \( X > S \) reading (for any \( X \in \{S, O, D\} \)) is ruled out if the relevant subjective probability measure is provided by a parameter of evaluation that is independent of the world of evaluation, as we are assuming. This assumption is borrowed from previous work, especially Yalcin 2007, and could of course be discarded by relaxing this assumption or by giving a more complex theory of embeddings.

On the other hand, we might find that \( S \)-readings are not totally unembeddable: \( S > S \) readings might occur when an agent is genuinely uncertain about her own beliefs, due to psychological limitations (e.g., time pressure rendering it difficult to reason through a problem, so that one’s own implicit beliefs have not yet emerged). This is in fact what I expect to find: when embedded \( S \)-readings occur, they should be in rare cases where it is sensible for someone to express uncertainty about her own subjective uncertainty.

In this section I will evaluate these hypotheses in a rough-and-ready way against examples of nested epistemics. Except where noted explicitly, the examples are drawn from the NYT-Eng section of the Gigaword corpus, comprising 1.4 million words of New York Times newswire. The NYT examples were selected by retrieving all sentences in which two or more of the following items occurred within the same clause (as delimited by the punctuation marks [!,;,.?]).

probable, probably, improbable, likely, certain, sure, definitely, possible, possibly, impossible, must, might, may, maybe, perhaps.
To keep the study to a manageable size I did not include many relevant items such as confident, clear, clearly, dubious, think, believe, know, doubt, chance, likelihood, probability, possibility, confidence.

This search technique returned a large number of hits, most of them irrelevant: for example, examples of must with a deontic interpretation, certain in expressions such as a certain individual, and examples in which the two modals were not in a scopal relation (e.g. probable and maybe certain). However, within this collection I was able to find a substantial number of examples which do appear to be genuine nested epistemics.

I selected for discussion here examples that had a relatively clear intended meaning, and which were relevant to a theoretically interesting point. All interpretations are my own. A controlled evaluation using trained, hypothesis-naïve annotators would of course be preferable, but this will have to wait for future work.

5.1 Subjective uncertainty about casually relevant chance factors ($S > O$)

It’s relatively easy to find examples in which an author expresses subjective uncertainty about chance factors. For example, (7) appears to have this character.

(7) Nuclear war is as likely, and perhaps more likely, in our time than at any point during the Cold War because superpower constraints are much weaker ... (perhaps > more likely)

The message is that complex, interacting geopolitical forces have conspired such that the objective probability of nuclear war is (definitely) at least as high as at any point during the cold war—and that the information available to the author renders it possible that this objective probability is even higher. So, perhaps more likely has a $S > O$ reading.

Here is another, somewhat more complex example.

(8) Despite starting 29th in today’s Samsung/Radio Shack 500, Gordon had a spring in his step on Saturday at Texas Motor Speedway after posting the fourth-fastest lap during the final practice session. “We’re probably the dark horse for sure,” Gordon said. “Our cars have been really, really fast lately. ... Maybe Sunday will be our day.” (for sure > probably)

I believe that the intended scopal order here is for sure > probably. Assuming this, Gordon’s message seems to be that he is subjectively certain that objective features of the world (his team’s abilities, other teams’ performances) will be arranged so that his team probably performs above expectations on Sunday.

(9) To build a large database of people who are interested in participating as links in the chain, Watts’ team has set up a Web site ... Participants will receive e-mail messages asking them to help locate a target person by forwarding e-mail to a personal contact who might be likely to know the target.

As I understand this example, Watts’ team is asking participants to identify people who might (given their information about the individuals in question) be likely (given objective facts about that individual, such as their location and activities) to know the target.
The following is a relatively clear case, where subjective *definitely* embeds the epistemic expression *much more likely*. The latter is explicitly justified in terms of current conditions in the global economy—a radically indeterminate macrostructure if there ever was one.

(10) “You can be guaranteed of rocketing commodity prices if and when the global economy starts working together,” said Bond. “We’ve definitely got conditions now where commodity price squeezes are much more likely than they were in the past.” (definitely > likely)

(11) The new federal law makes most legal immigrants ineligible for Supplemental Security Income ... All those people might be likely applicants for Home Relief.

If I understand correctly, the intended interpretation of (11) is that, for each newly ineligible immigrant $x$, it is subjectively possible that $x$ would be likely (in light of objective facts about the individual and the application process) to apply for Home Relief. (As an aside, (11) is a counter-example to von Fintel & Iatridou’s (2003) “epistemic containment” constraint, which forbids quantifiers from taking scope directly above an epistemic modal. See Swanson 2010 for additional counter-examples.)

5.2 Subjective uncertainty about distributional probabilities ($S > D$)

The only good candidate of a potential $S > D$ example that I have located is (12):

(12) We do have an enemy, and it is definitely possible that what you consider a “word from God” could in fact be demonic and be “a word from satan.” (web)

An $S > O$ reading of this example is not very plausible, since it is presumably a determinate present fact whether a given communication is from God or from Satan (or indeed from neither). If so, only $S > S$ and $S > D$ would make sense here. My interpretation of this example goes as follows: the author is expressing maximal confidence in the claim that Satan has the ability to deceive people into thinking that his words come from God. In other words, the author is confident that—among the class of communications that his audience consider to be “words from God”—a non-zero proportion of these are in fact “words from Satan”. If so, the quantity that he is maximally confident of is a distributional probability.

5.3 $D > S$?

I have found no corpus examples that suggest a $D > S$ interpretation. An invented example along the following lines was recently suggested to me:

(13) He is a Republican; so, it’s unlikely that he is probably a pacifist.

To the extent that the example makes sense, it suggests distributional reasoning: if you sample an individual at random from the class of Republicans, it is unlikely that the individual sampled will be one who is probably a pacifist. We might imagine, after sampling an individual, being forced to infer from observables whether the individual is a pacifist. In any case, if example (13) is
intelligible it may provide another argument that subjective readings of probability expressions can be embedded under other epistemic operators.

5.4 **Multiple layers of objective uncertainty** \((O > O)\)

Most of the examples that follow could also be interpreted as \(S > O\). The best \(O > O\) candidates seem to involve well-informed speculation about the future, which (as we noted above) people seem to treat as open and in-principle unknowable.

(14) NBA Commissioner Adam Silver said Wednesday night that it’s “probably unlikely” for Orlando to host the 2017 NBA All-Star Game because the city hosted the exhibition just several years ago, in 2012. ... Silver answered that Orlando’s chances are slim “only because the game was here in 2012 and there are some other teams that haven’t had an All-Star Game in a long time ... But having said that, this was a great experience down here in 2012, and I’m sure we won’t be waiting to go to 29 other teams before we come back here.” *(Orlando Sentinel, 2/25/15)*

Here Silver expresses uncertainty about the future value of a set of variables—other teams’ levels of interest—which are known to be causally relevant to Orlando’s chances of getting the Game. (This choice is ultimately under Silver’s control.) It’s hard to be sure that the “probably” here is of the objective sort, but if it is, this is an \(O > O\) example.

(15) Tom McGowan, the town planner, suggested that if critical sites needed protection that could not be provided in any other way, “a very focused proposal might be likely to receive favor.” He cited such vulnerable open areas as Routes 118 and 254, both entry roads into the village center.

We might think to treat (15) as \(D > S\), but I think that this would be incorrect: McGowan is not discussing the prospects of a proposal randomly selected from a known class, but rather noting that, if a focused proposal is written, it might (depending on its characteristics) be likely (depending on the planning commission’s internal deliberations) to receive favor. If so, this can be read as an \(S > S\) example.

5.5 **Genuine higher-order uncertainty** \((S > S)\)

Many theorists have wished to ban true higher-order uncertainty, of the \(S > S\) variety, for philosophical or computational reasons. It’s true that allowing higher-order uncertainty makes the semantics much more complicated, and that it conflicts with certain theoretical positions about the nature of belief or introspection. Still, it may be a real thing. Here are some examples that could be analyzed as \(S > S\) readings.

(16) “I believe in a ‘yes’ victory, but I’m definitely not certain,” said Prime Minister Goran Persson ...

I don’t see any alternative except to understand this example as an expression of Persson’s higher-order certainty about the non-maximal value of his own subjective degree of belief. (Aside: this
example illustrates the “weakness of belief”: Persson believes, but is (definitely!) not certain. See Hawthorne, Rothschild & Spectre 2016 for discussion.) Here is another similar example:

(17) This book is an odyssey, a journey up through the mists of time from the remote past. It explores what must perhaps be the most fundamental of all questions - who we are. (web)

As I understand it, the author is not at all sure about what the most fundamental of all questions is—but he is virtually certain that it might be “who we are”. So, this looks like another $S > S$ reading.

Once we have opened this can of worms, further candidates start to appear.

(18) Mr. Nolan must certainly be rubbing his conspiratorial hands with glee as he hears that the arrival of Primark has meant that other stores are now looking to make a new home in Truro. (web)

This example would typically be interpreted as modal concord. This is not an unreasonable analysis, but note that examples (16) and (17) can’t be treated in this way since the modals are not matched in strength. Perhaps some constructions that could plausibly be used to express concord can also yield to a nested analysis. I am tempted to read (18) as expressing higher-order confidence, either about the author’s own certainty or about the appropriate degree of certainty for conversational partners to adopt in light of the evidence available.

I don’t know if any of these examples will hold up in the end: perhaps they can all be explained away in some other fashion. But it may also be that authors sometimes wish to express (un)certainty about their own beliefs, about how their beliefs will stabilize upon reflection, or about what the appropriate degree of belief is for people with access to a given body of information.

6 Distributional considerations

There appear to be significant gaps in the distribution of nested modals in the Gigaword-NYT corpus. For example, none of the following occur within the same clause within the corpus.

- definitely ... may/(im)probable
- impossible ... definitely/maybe
- perhaps ... definitely
- possibly ... improbable/probable
- improbable ... probably/certain/definitely/may/must/perhaps/possibly/sure

In a corpus as large as this one, where many combinations are attested by numerous examples, these gaps are striking. Some of them may be explicable in terms of Ülkümen et al.’s (2015) observations about preferred interpretations of modals: as I mentioned above, they point out that certain, sure, confident, and expressions derived from them tend to deal with subjective uncertainty, while likely, likelihood, chance, probability and their derivatives prefer an external (objective or perhaps distributional) interpretation. If such preferences do exist, then we might be able to explain certain gaps by pointing to the semantic ill-formedness, or pragmatic infelicity, of the preferred readings of certain combinations.
7 Embedding epistemics simple-mindedly

One of the goals of this paper is to consider the prospects of a simple-minded analysis of nested epistemics: probability expressions of all stripes embed propositions (sets of possible worlds), and anything that can serve as their direct argument must pick out a set of possible worlds. If we start with Yalcin’s (2007) hypothesis that subjective probabilities come into the composition via a parameter that is independent of the world of evaluation, we expect that subjective probability expressions should not embed non-trivially, but objective and distributional probability expressions—being world-bound—should.

The examples tentatively analyzed as $S > S$ in section 5.5 do not yet refute this position, but they do problematize it somewhat. If we wish to hold onto a simple theory of embeddings, we may have to “objectivize” subjective probabilities in some cases, treating them as features of the world about which we can be uncertain. There does not seem to be any technical barrier to doing so, though there might be philosophical reasons to try to find a way around it. Alternatively, we could frame a more complicated theory of embeddings, as Moss (2015) does. While this theory does many things well, it has certain empirical drawbacks, and it is definitely much more complex and radical in its semantic and philosophical implications than the simple-minded approach suggested here.

This bring us back to the vacuity issue in section 1. Do multiple epistemics reduce to the innermost in English, as they do in S5? On face, the answer is a clear “no”: the nested examples that we have analyzed have complex, subtle interpretations that definitely are not equivalent to the same sentences with the outermost modal deleted. However, the English data are not relevant to this question if there is a subtle ambiguity in the interpretation of the modal expressions, as I argued that there is in many cases. A convincing test of the empirical claim would have to involve $S > S$, $O > O$, or $D > D$ interpretations. $D > D$ readings do not seem to occur among the examples that I have found. (I am unsure what they would even mean). Various examples of $S > S$ and $O > O$ readings did occur, but these could probably be re-analyzed in some other way by a dedicated proponent of the S5 reduction property.

This conclusion is somewhat unsatisfying, since we don’t know for sure yet whether reduction holds. Reduction does have two nice features that might make it worth pursuing, though. First, if it is a simple-minded theory of embeddings is in the clear—at least so far, and in terms of the basic data. Second, it means that we will have to continue doing what Lauri taught us many years ago: continue to take inspiration from logic and philosophy, but evaluate every idea inspired from these sources carefully against empirical data. The only major methodological development to report is that, through the new probabilistic orientation in modal semantics, we can now draw on psychological work of reasoning and choice that take inspiration from probabilistic models.

References


Goodhue, Daniel. 2018. *Must p is felicitous only if p is not known*. To appear in *Semantics & Pragmatics*.


15