Gradable epistemic modals, probability, and scale structure[*]

Daniel Lassiter
New York University

Abstract The epistemic modals *possible, probable, likely*, and *certain* require a semantics which explains their behavior both as modal operators and as gradable adjectives. An analysis of these items in terms of Kennedy & McNally's (2005) theory of gradability suggests that they are associated with a single, fully closed scale of possibility. An implementation using the standard theory of modality due to Kratzer is shown to make incorrect predictions in several domains. However, identifying the scale of possibility with standard numerical probability explains the facts about gradability and avoids the undesirable predictions of Kratzer's theory.

Keywords: gradability, comparison, epistemic modals, probability

1 Gradable modals

Most discussion of the semantics of English modals has focused on the meanings of modal auxiliaries such as *must*, *should*, and *can*, illustrated in (1).

(1)   a. Harry <u>should</u> be in Sacramento by now.

b. My brother <u>can</u> bench press 250 pounds.

c. All cameras <u>must</u> be checked at the door.

However, English also has a substantial number of adjectival modals such as *likely*, *obligatory*, *able*, and *evident*. These items are readily **gradable**, just like the large and well-studied class of gradable adjectives. Some examples are given in (2a)-(2d).

---

(2)   a. Degree Modification:
         It is very likely that Jorge will win the race.

       b. Comparison:
         It is more likely that Jorge will win the race than it is that Sue will win.

       c. Degree questions:
         How likely is it that Jorge will win the race?

       d. Explicit Degree Quantification:
         It is 95% certain that Jorge will win the race.

As Yalcin (2007) and Portner (2009) note, it should be possible given these facts to use existing theories of the semantics of gradable adjectives such as Kennedy & McNally 2005 and Kennedy 2007 to constrain our theory of epistemic modality. Steps in this direction have been taken by Portner (2009) and Yalcin (2010b).

The present paper is a contribution to this small but growing literature, focusing on the gradable epistemic modals (**GEM**s) *possible*, *probable*, *likely*, and *certain*. After introducing the scale-based theory of Kennedy & McNally (2005), I analyze the behavior of these GEMs with respect to various tests for adjective type and scale structure, showing that their differences in behavior follow from their differing locations on a single scale of possibility, one that has both top and bottom elements.

Since the standard theory of modality (Kratzer 1981, 1991) is built around a notion of comparative possibility, we might expect this to form the basis of the scales in question. However, I argue this approach makes incorrect predictions about ratio and proportional modifiers and about the interaction of disjunction with equatives. An alternative is to suppose with Yalcin (2010b) that GEMs are associated with the scale of standard numerical probabilities. I show that this assumption explains the behavior of GEMs with respect to tests for scale structure, and avoids Kratzer's problems with quantitative expressions of likelihood and with disjunction.

2   Kennedy & McNally's theory of gradability

2.1   Measure functions and degree modifiers

According to Kennedy & McNally (2005) (K&M, see also Kennedy 2007), gradable adjectives denote measure functions, i.e. functions from individuals to points on a scale. The nature of the scale is (at least partially) determined by the lexical semantics of the adjective in question. So, for example, *tall* denotes a function from an object to its height.[1] Formally, a scale is given by a triple $\langle D, <, \delta \rangle$, where D is

---

1 Alternatively we might take adjectives to denote functions from individuals to intervals on a scale (Schwarzschild & Wilkinson 2002; Kennedy 2001), or relations between degrees and individuals (Heim 2001). The choice between these options is not crucial for current purposes.

a set of degrees, $\prec$ is a total ordering of D, and δ is the *dimension* of the adjective (e.g., in the case of *tall*, the dimension of height). The meaning of *tall* is as in (3):

(3) $[\![tall]\!] = \lambda x.\iota d.\ x$'s height $= d$

When an adjective has an explicit degree modifier, the adjective is an argument of the modifier, as in (4).

(4)   a. $[\![closed]\!] = \lambda x.\iota d.\ x$'s degree of closure $= d$
       b. $[\![completely]\!] = \lambda A.\lambda x.A(x) = \mathbf{max}(D_A)$
       c. $[\![completely\ closed]\!] = \lambda x.[\iota d\!: [\![closed]\!](x) = d] = \mathbf{max}(D_{closed})$

Hence *The door is completely closed* means that the degree of closure of the door is the maximal degree in D_{closed}. This predicts immediately that we cannot use *completely* on its degree-modifier reading with an adjective associated with a set of degrees with no maximal element — a correct prediction, as we will see.

Gradable adjectives in this theory denote functions from individuals to degrees — not the right type for a predicate, of course. K&M argue that when an adjective is used in the positive form, it is converted into a predicate by a null morpheme **pos**, which supplies a standard of comparison $\mathbf{s}_A$.

(5)   a. $[\![\mathbf{pos}]\!] = \lambda A.\lambda x.A(x) > \mathbf{s}_A$
       b. $[\![tall]\!] = \lambda x.\iota d.\ x\text{'s height} = d$
       c. $[\![\mathbf{pos}\ tall]\!] = \lambda x.[\iota d.\ x\text{'s height} = d] > \mathbf{s}_{tall}$

How $\mathbf{s}_A$ is determined is a complex issue which we won't address here; see Kennedy & McNally 2005 and Kennedy 2007 for much discussion.

2.2   Parameters of variation: Scale structure and adjective type

One of the most important parameters of variation among gradable adjectives, according to K&M, is the structure of the domain of degrees D. They distinguish four scale types: **totally open**, **totally closed**, **upper bounded**, and **lower bounded**.
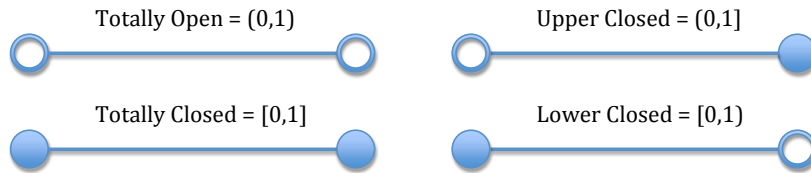

Fig. 1. Scale Types.

A totally open scale has no top or bottom element, while a totally closed scale has both. The other two scale types receive the obvious interpretations.

A closely related parameter of variation is **adjective type**. K&M discuss three types, the **maximum-standard**, **minimum-standard**, and **relative-standard** adjectives. Maximum-standard adjectives like *full*, in the positive form, pick out the top portion of a scale: for example, a glass is full just in case it cannot contain any more liquid. Maximum-standard adjectives must be associated with a scale with a top element, i.e. an upper-closed or fully-closed scale. Minimum-standard adjectives like *bent* pay attention to the lower portion of a scale: something counts as bent if its degree of bend is non-zero (i.e., if it is not straight). This requires that minimum-standard adjectives are associated with a lower-closed or fully-closed scale. Finally, relative-standard adjectives like *tall* have a value that is more context-dependent, but generally falls somewhere in the mid-range of the scale. Typically they are associated with scales without a top element (as with *tall*), although, I will argue, this is not always the case (cf. discussion of *inexpensive* and *lightweight* in §3.5).

2.3   Tests for adjective type and scale structure

This subsection presents seven tests for adjective type and scale structure. Except where noted, tests are from Kennedy & McNally 2005 or Kennedy 2007.

1. *Completely*-**modification.** *Completely* is a polysemous modifier, but one common function is as a maximizer. K&M argue that, if an adjective can be modified by *completely* with a "maximum" interpretation, it is a maximum-standard adjective, and thus denotes a function whose range is an upper closed scale. This explains the contrast in (6), where *bent* and *tall* are unacceptable with degree-modifying *completely* because their scales do not have a maximum element.

(6)   a. The room is completely full.

        b. # This basketball player is completely tall.

        c. # The rod is completely bent.

2. *Slightly*-**modification.** According to K&M, *x is slightly A* is true just in case $[\![A]\!](x)$ differs by a small amount from the minimum degree on the scale associated with A. Thus if an adjective can be modified by *slightly*, it is associated with a scale with a bottom element, and, in general, is a minimum-standard adjective.

(7)   a. The rod is slightly bent.

        b. # This player is slightly tall.

        c. # The room is slightly full.

3. "A but could be A-er": Minimum-standard and relative-standard adjectives allow the construction in (8), but maximum-standard adjectives do not.

(8)   a. The rod is bent, but it could be more bent.

b. This basketball player is tall, but he could be taller.

c. # The room is full, but it could be fuller.

4. Negation entails antonym? The negation of a minimum- or maximum-standard adjective entails that its antonym holds. Relative-standard adjectives have a "zone of indifference" (Sapir 1944), so that it is possible to be neither *A* nor *not-A*.

(9) a. The rod is not bent. $\models$ The rod is straight.

b. The rod is not straight. $\models$ The rod is bent.

c. This player is not tall. $\not\models$ This player is short.

5. Type of antonym. The defining characteristic of antonymous pairs of adjectives is that *x is more A_{pos} than y* is true iff *y is more A_{neg} than x* is true. I will assume that antonymy simply involves reversing the direction of the ordering relation.[2] It follows that the maximum element of a scale, if there is one, is always the minimum element of its antonym, and vice versa (Rotstein & Winter 2004). As a result, maximum-standard adjectives have minimum-standard antonyms and vice versa:

(10) a. This neighborhood is completely/#slightly safe.

b. This neighborhood is slightly/#completely dangerous.

(11) a. The rod is slightly/#completely bent.

b. The rod is completely/#slightly straight.

However, the antonym of a relative adjective is also a relative adjective.

(12) a. My car was #completely/#slightly expensive.

b. My car was #completely/#slightly cheap.

6. *Almost*. If an object is *almost A*, then it fails to be *A* by a small margin. Rotstein & Winter (2004) show that *almost A* is acceptable only with with maximum-standard adjectives *A*, and entails that the antonym of *A* with *slightly* holds of *x*:

(13) a. This area is almost safe. $\models$ This area is slightly dangerous.

b. #This player is almost tall.

c. #This rod is almost bent.

7. Proportional Modification. Proportional modifiers like *half*, *70%*, and *mostly* quantify the difference between a degree's location on the scale relative to both the maximum and minimum points.

(14) a. $[\![half]\!] = \lambda A.\lambda x.A(x) = \frac{1}{2} \times [max(D_A) - min(D_A)]$

---

2 The truth is more complicated than this, though: see e.g. Kennedy 2001 and Sassoon 2010.

b. $[\![70\%]\!] = \lambda A.\lambda x.A(x) = .7 \times [max(D_A) - min(D_A)]$

These modifiers occur only with adjectives such as *full* which are associated with a fully closed scale, since they are undefined if either of these points does not exist.

(15)   a. The glass is half full/empty.

b. # My cousin is half tall/short.

c. # This road is half dangerous/safe.

In most cases proportional modifiers occur with maximum-standard adjectives. However, I will show in §3.5 that they also occur in some cases with minimum- and relative-standard adjectives which are located on fully closed scales.

3   The scale structure of gradable epistemic modals

Our first task is to figure out where each of our GEMs falls in K&M's typology, and what the nature of the scale or scales in question is. Seth Yalcin, in unpublished work[3], has suggested that *possible* and *probable* are **probability operators**, i.e. that they map propositions to the range of standard probabilities $[0,1]$. This predicts that at least these two GEMs should be associated with a fully closed scale. On the other hand, Portner (2009) has claimed that, since *possible* and *probable* accept different degree modifiers, this shows that they are associated with different scales. Portner suggests that this approach is more congenial to the standard (non-probabilistic, non-quantitative) Kratzerian semantics for modality.

In this section I re-examine the evidence for the scale structure of GEMs systematically. Data regarding degree modification, as well as entailments and scalar implicatures, suggest that the GEMs *possible, probable, likely,* and *certain* are located on the same scale, and that this scale is fully closed. I also address the theoretical motivation for positing different scales based on degree modification data and argue that it is not compelling. All of this does not, of course, decide the question of whether probability is implicated, the subject of the following sections.

3.1   *Possible*

On all of our tests, *possible* behaves as a minimum-standard adjective.

(16)   a. #It is completely possible that the Jets will win.

b. It is slightly possible that the Jets will win.

3 Presented at *Sinn und Bedeutung* in 2006 and at NYU in 2009; cf. also the recent Yalcin 2010b.

 c. It is possible that the Jets will win, but it could be more possible.[4]

 d. It is not possible that the Jets will win. ⊨ It is impossible that they will.

 e. It is completely/#slightly impossible that the Jets will win.

 f. #It is almost possible that the Jets will win.

 g. #It is half possible that the Jets will win.

Since *possible* is a minimum-standard adjective, we expect, to a first approximation, the following truth-conditions:

(17) ϕ *is possible* is true iff ϕ's degree of possibility is not $\mathbf{min}(D_{poss})$.

And, of course, this means that the scale of possibility must have a minimum element.

3.2 Scale-mates, entailments, and implicatures

As Portner (2009) notes, *possible* and *probable* do not accept the same modifiers:[5]

(18) a. It is slightly possible that the Jets will win.

 b. # It is slightly probable that the Jets will win.

Portner concludes from this fact that *possible* and *probable* do not occupy the same scale. However, this does not necessarily follow: different adjectives occupying the same scale but accepting different modifiers is nothing unusual.

(19) The room is slightly occupied/#full.

What (18) does clearly show is that *possible* and *probable* have different ADJECTIVE TYPES: *probable* must be one of the types that do not accept *slightly*-modification, i.e. relative- or maximum-standard. Furthermore, *possible* and *probable* show entailment relations and implicatures that are straightforwardly explained if *probable* occupies a higher position on the same scale as *probable*.

(20) a. It is probable that we will win. ⊨ It is possible that we will win.

 b. It is not possible that we will win. ⊨ It is not probable that we will win.

This is the usual behavior of scalar items like *warm* and *hot*: *It is hot* entails *It is warm*, and (metalinguistic negation aside) *It is not warm* entails *It is not hot*. *Possible* and *probable* also behave like *warm* and *hot* with respect to scalar implicatures:

---

4 Many speakers accept the comparative *more possible*, though some express discomfort, preferring *more likely* even for small values. I do not know the source of this preference, but it does not seem to be grammatical in nature: *more possible* is robustly attested in corpora.

5 Portner actually uses different examples to make this point. I use *slightly* because the examples that Portner marks as unacceptable (*completely/entirely probable*, *extremely/more possible*) are actually well-attested in corpora.

(21)   a. It is possible that we will win. ⇝ It is not probable that we will win.

         b. It is not probable that we will win. ⇝ It is possible that we will win.

The simplest explanation of these facts is that *possible* and *probable* occupy different regions of a single scale of possibility, just as *warm* and *hot* occupy different regions of the temperature scale.

3.3 *Probable* and *likely*

Probable and *likely* appear to be synonymous, although there are syntactic and register differences between them. Since they share a scale with *possible*, but are not minimum-standard, we can now ask whether they are relative- or maximum-standard. Almost without exception, they behave as relative-standard adjectives:

(22)   a. #It is completely likely/probable that the Jets will win this year.[6]

         b. #It is slightly likely/probable that the Jets will win.

         c. It is likely that the Jets will win, but it could be more likely.

         d. It is not likely that the Jets will win, but it is not unlikely either.

         e. It is #completely/#slightly unlikely/improbable that the Jets will win.

         f. #It is almost likely/probable that the Jets will win.

         g. #It is half likely/probable that the Jets will win.

The exception is that, in contrast with the relative adjective *tall*, *likely* and *probable* occur with the proportional modifier *n%* quite often in corpora:

(23)   It is 70% likely that the Jets will win.

An explanation of this difference will emerge in the following subsection: *likely*, unlike *tall*, is associated with a scale with a maximum element.
    All of this suggests the following rough truth-conditions for *possible* and *likely*:

(24) ϕ *is likely/probable* is true iff ϕ's degree of possibility is greater than s_{poss} (the contextually determined standard; cf. (5)).

3.4 *Certain*

Similar reasoning about entailment and implicature suggests that *certain* occupies the same scale as *probable* and *likely*:

(25)   a. It is certain that we will win. ⊨ It is likely that we will win.

---

6 (22a) is acceptable when *completely* indicates a correction, but not when it is a degree modifier.

b. It is not likely that we will win. ⊨ It is not certain that we will win.

c. It is likely that we will win. ⤳ It is not certain that we will win.

d. It is not certain that we will win. ⤳ It is likely that we will win.[7]

On all of our tests, *certain* behaves as a maximum-standard adjective:

(26)   a. It is completely certain that the Jets will win.

b. #It is slightly certain that the Jets will win.

c. #It is certain that the Jets will win, but it could be more certain.

d. It is not certain that the Jets will win. ⊨ It is (at least a little bit) uncertain.

e. It is slightly uncertain that the Jets will win.

f. It is almost certain that the Jets will win.

g. It is almost certain that the Jets will win. ⊨ It is slightly uncertain.

h. It is half/95% certain that the Jets will win.

These facts suggest the following rough truth-conditions:

(27) ϕ *is certain* is true iff ϕ's degree of possibility is $\mathbf{max}(D_{poss})$.

And, of course, it follows that the scale of possibility must have a maximum element.

3.5   Failure to Maximize with *completely*

Portner (2009) points out an interesting problem with the claim that the GEMs occupy a single scale: it seems to predict that *completely possible* and *completely probable* should have the meanings in (28).

(28)   a. ϕ *is completely possible* is true iff ϕ's degree of possibility = $\mathbf{max}(D_{poss})$

b. ϕ *is completely probable* is true iff ϕ's degree of possibility = $\mathbf{max}(D_{poss})$

That is, both should mean the same thing as *certain*. Since this is evidently false, Portner concludes that these adjectives do not occupy an upper-closed scale.

The problem in (28) is real, but it is a general problem for K&M's theory of gradability, not a special fact about *possible* and *probable*. In fact, Kennedy (2007) points out effectively the same issue with respect to (29a) and (29b), which, on intuitive assumptions about the nature of the scales in question, should be equivalent; (30) is another example.

(29)   a. This pizza is completely inexpensive.

---

7 If you feel squeamish about (25d), try substituting *It is possible that we will win*, and the reasoning goes through. To the extent that the implicature in (25d) is weak, this is matched by other scalar items: e.g., the implicature from *It is not hot* to *It is warm* is not especially robust.

      b. This pizza is free.

(30)   a. On that planet, you would be completely lightweight.

       b. On that planet, you would be weightless.

So the conclusion that Portner draws from (28) is at least not obviously valid: if (28) shows that *possible* and *probable* do not have upper-closed scales, then (29)-(30) show, against intuition, that *inexpensive* and *lightweight* do not either.

    Do we want to conclude that the scales that *inexpensive* and *lightweight* are associated with exclude the values that intuitively constitute their maxima (zero cost and zero weight respectively)? Kennedy (2007) endorses this claim: although "naive intuition suggests that the COST scale should have a minimal value representing complete lack of cost", (29) shows that it does not: "Instead, the property of zero cost is named by the non-gradable adjective *free*".

    There are several problems with this analysis. First, it's just not true that *free* is a non-gradable adjective — for example, *completely/perfectly free* and *almost/nearly free* are acceptable, unlike the same modifiers with a truly non-gradable adjective like *geological*. These and other tests indicate that *free* is a maximum-standard adjective on an upper-closed scale, as expected. Further, since an object is *almost free* just in case its cost is small but non-zero, the scale in question must be the scale of costs. Similar facts hold for *weightless*.[8] So we cannot say that *free* and *inexpensive* occupy completely different scales. At best, we could stipulate that *inexpensive* does not have access to the top element of the COST scale, but this would seem to be a restatement of the facts with little explanatory value.

    An even more serious problem is that, even if this is the correct analysis of *free*, it encounters additional difficulty with GEMs, all of which accept the proportional modifier *n percent*. We saw this for *likely* already in (23). A few more examples:

(31)   a. I felt that if it was 80-90 percent possible that [the cancer] hadn't spread, I didn't want the hysterectomy. [9]

       b. [I]t's 80 percent likely that the iPhone will be coming to T-Mobile ... [10]

       c. [T]he IPCC ... said it was "very likely" or more than 90 percent probable that human activities ... had caused most of the warming in the past half century.[11]

---

8 The main peculiarity of *free* and *weightless* seems to be that they are awkward in the comparative. It is not clear to me why this should be, but, as Kennedy (2007: 34) notes, "not all modifiers co-occur with all adjectives for apparently idiosyncratic reasons".

9 http://www.suggestadoctor.com/doctor_118_paulshuchiang_lin.htm

10 Huffington Post, August 1, 2010 (http://www.huffingtonpost.com/2010/07/22/tmobile-iphone-in-2010-80_n_655459.html)

11 Reuters, November 17, 2007 (http://www.reuters.com/article/idUSGOR68185720071117)

Since proportional modifiers crucially make reference to the top and bottom elements of a scale, these data forced us to the conclusion that **all** of the GEMs have a fully closed scale. This means that we cannot explain their peculiarities with respect to *completely* by simply stipulating that their scale does not include a top element.

A more general issue is that the proposed solution amounts to gerrymandering scale structure in order to explain anomalies in the degree-modification tests. Much of the attractiveness of K&M's theory of scale structure is due to the fact that their tests typically associate adjectives with scales that map onto plausible ontological distinctions. For example, evidence that *full* has a fully closed scale fits naturally with the observation that there are upper and lower limits to how much liquid a container can hold. If the tests suggested instead that *full* is associated with a scale with no top or bottom element, I suspect that we would conclude that there is a problem with our interpretation of the tests that we have devised, rather than taking this as evidence that *full* has a counter-intuitive scale structure. I think that we should draw the same conclusion for *inexpensive* and, *mutatis mutandis*, for *possible* and *probable*: the problem is not with the scales, but with the interpretation of the tests.

The right lesson to draw from (28)-(30) and similar data, then, is that tests for scale structure constitute one-way implications. If *completely A* denotes a maximal value, then *A* has an upper closed scale; but it is not necessarily the case that, if *A* has an upper closed scale, then *completely A* will denote a maximal value. Other factors may intervene to complicate matters. In the case of (28b)-(30), the intervening factors are relatively clear: in each case, a relative adjective fails to maximize with *completely* when there is a maximum-standard adjective (*free, weightless, certain*) occupying the endpoint of the scale. Likewise, in (28a) the minimum-standard adjective *possible* fails to maximize in the same situation. We could interpret these facts as supporting any of the following generalizations:

(32) a. Relative- and minimum-standard adjectives cannot be modified by degree-maximizing *completely* even if their scale is upper closed.

b. Relative- and minimum-standard adjectives whose scale is upper closed cannot be modified by degree-maximizing *completely* if there is a maximum-standard adjective on the same scale.

c. Relative- and minimum-standard adjectives only occur on upper-closed scales if there is a maximum-standard adjective on the same scale (which then blocks degree-modifying *completely*).

The data do not yet allows us to choose between these options, but it seems that one of the generalizations in (32) is needed to explain (28)-(30). However these data are explained, though, it does not seem that (28) motivates placing the various GEMs on separate scales. And if we were to do so, we would lose the explanation of the various entailments and implicatures noted above, as well as the intuitive ontological

motivations of the theory with respect to adjectives like *inexpensive* and *lightweight*.

3.6 The big picture

The overall picture that emerges is this: GEMs are functions from propositions to points on a single scale of possibility. This scale is fully closed and contains the minimum-standard adjective *possible*, the relative-standard adjectives *probable* and *likely*, and the maximum-standard adjective *certain*. These behave as expected with respect to degree modification, entailments, and implicatures, with one peculiarity: all of them, even the minimum- and relative-standard ones, accept certain proportional modifiers. This is explained by the fact that their scale is fully closed.[12]

4 Semantics for epistemic modals: The standard theory

The standard theory of modality among linguists is due to Angelika Kratzer (Kratzer 1981, 1991). This theory is built around a notion of **comparative possibility**, making it a very promising candidate to supply a degree semantics for GEMs. This section outlines such an approach, and then notes two serious problems that arise.

4.1 The basic theory

Kratzer's theory of modality is built around a binary relation on worlds $\geqslant$ which holds of two worlds u, v just in case u is "at least as good a possibility" as v in a way to be made precise.[13] $\geqslant$ is interpreted relative to a modal base $\mathbf{f}$ and an ordering source $\mathbf{g}$. $\mathbf{f}$ is a function which, given a world, returns a set of propositions that are relevant to the evaluation of the modal expression. In the case of epistemic modality, the modal base is the set of propositions known to or believed by the speaker (or whoever else the contextually appropriate people are). $\mathbf{g}$ is a function which, applied to a world w, returns a set of propositions which induces a preorder (quasi-order) over the worlds in $\bigcap \mathbf{f}(w)$ according to which of the propositions in $\mathbf{g}(w)$ they satisfy. The ordering is determined by the rule in (33):

(33) $\forall u, v \in \bigcap \mathbf{f}(w) \colon u \geqslant_{\mathbf{g}(w)} v$ iff $\{p \colon p \in \mathbf{g}(w) \wedge v \in p\} \subseteq \{p \colon p \in \mathbf{g}(w) \wedge u \in p\}$

---

12 Readers familiar with the Kennedy-McNally theory of gradability may have noted that the claim that the four GEMs we are interested in all occupy the same scale is not compatible with certain aspects of Kennedy 2007. In particular, Kennedy's attempt to derive adjective type from scale structure ("Interpretive Economy") rather than treating it as a lexical feature of adjectives (as K&M do) is not compatible with the proposal made here — there are simply too many adjectives on the scale of possibility. However, the present proposal is compatible with Potts's (2008) reinterpretation of Interpretive Economy as a historical/evolutionary pressure.

13 I use $\geqslant$, etc. where Kratzer (1981) uses $\leqslant$, etc.

That is, world u is at least as good as world v relative to $\mathbf{g}(w)$ iff u is an element of every proposition in $\mathbf{g}(w)$ that v is. u is strictly better, $u > v$, iff $u \succcurlyeq_{\mathbf{g}(w)} v$ and $\neg(v \succcurlyeq_{\mathbf{g}(w)} u)$. (From here on I will mostly suppress reference to $\mathbf{g}(w)$, since variation in the ordering source will not be important for our purposes.)

Comparative possibility can be extended to a relation on propositions as in (34) (Kratzer 1991; Halpern 1997), where $\phi \succcurlyeq^s \psi$ is read "ϕ is at least as likely as ψ".

(34) $\phi \succcurlyeq^s \psi$ iff $\forall w[w \in \psi \rightarrow \exists w'(w' \in \phi \wedge w' \succcurlyeq w)]$

In words, ϕ is at least as likely as ψ iff every ψ-world is weakly dominated by at least one ϕ-world. ϕ is <u>more</u> likely than ψ ($\phi >^s \psi$) iff $(\phi \succcurlyeq^s \psi) \wedge \neg(\psi \succcurlyeq^s \phi)$.

4.2   Extension to gradable epistemic modals

A minimal modification of Kratzer's theory to account for GEMs might run along the following lines. K&M require scales to be totally ordered. This is crucially **not** true of the reflexive, transitive relation $\succcurlyeq^s$, which would naturally generate a partially ordered (reflexive, transitive, antisymmetric) set of degrees. One approach might be to weaken K&M's theory to allow for partially ordered domains of degrees. I will follow a simpler approach by defining the class of **admissible epistemic possibility functions** μ that satisfy the requirements of K&M's theory and the results of section 3, leaving the choice of μ underdetermined.

(35)   Given a relation $\succcurlyeq^s$, an admissible epistemic possibility function is any function μ from propositions to $[\mathbf{min}(D_{poss}), \mathbf{max}(D_{poss})]$ where

a. μ does not violate the ordering $\succcurlyeq^s$: if $\phi \succcurlyeq^s \psi$, then $\mu(\phi) \geq \mu(\psi)$

b. $\mu(\varnothing) = \mathbf{min}(D_{poss})$ and $\mu(W) = \mathbf{max}(D_{poss})$

Any μ satisfying (35) will preserve the truth of all comparative possibility statements. However, like $\succcurlyeq^s$, μ contains no meaningful **quantitative** information: for example, the ratio of $\mu(\phi)$ to $\mu(\psi)$, or their difference, is not stable for different choices of μ satisfying (35) (except in the special case where $\phi \approx \psi$).

4.3   Problems with the degree-based extension

The degree-based variant of Kratzer's theory just outlined accounts for locutions of the form "ϕ is more likely than ψ", "ϕ is at least as likely as ψ", and so on. However, there are several problems with the approach due to the lack of meaningful quantitative information contained in $\succcurlyeq^s$ and the class of epistemic possibility functions it defines. I will mention two here.

The first set of problems to note involve degree modifiers which require quantitative comparisons of likelihood. For example, all of the expression-types in (36)

involve quantitative comparison, either between two propositions or between a proposition and the maximum and minimum of the scale.

(36)   a. ϕ is twice as likely as ψ.

b. It is half certain that ϕ.

c. It is 95% certain that ϕ.

Given a comparative possibility relation $\geqslant$, for any admissible μ' that would make the sentences in (36) true, we can choose a μ'' that would make the same sentences false while still respecting the underlying comparative possibility relation. (In the jargon of measurement theory (Krantz, Luce, Suppes & Tversky 1971), quantitative statements like those in (36) are formally MEANINGLESS because $\geqslant^s$ is even weaker than an ordinal scale, so that all monotone increasing transformations of μ are admissible, and many more.)

A second problem is that Kratzer's theory, both in its original form and in the extension given here, predicts the validity of a class of inferences involving disjunction which are clearly invalid.

(37)   **The Disjunctive Inference**

a. ϕ is at least as likely as ψ.

b. ϕ is at least as likely as χ.

c. $\therefore$ ϕ is at least as likely as $(\psi \vee \chi)$.

Proof of (37). (37a) means that, for every ψ-world v, there is a ϕ-world u such that $u \geqslant v$. Likewise, (37b) means that for every χ-world v' there is a ϕ-world u' such that $u' \geqslant v'$. Let w be an arbitrary world in $\psi \vee \chi$. Case 1: $w \in \psi$. Then there is a ϕ-world w' such that $w' \geqslant w$, namely u. Case 2: $w \in \chi$. Then there is a ϕ-world w' such that $w' \geqslant w$, namely u'. Since w was arbitrary, we conclude that for every $w \in (\psi \vee \chi)$, there is a $w' \in \phi$ such that $w' \geqslant w$; thus, by definition (34), (37c) holds.[14]

The problem is that the Disjunctive Inference is intuitively invalid; this is particularly clear when (37) is applied repeatedly. Suppose someone tells you:

(38)   "For any team you like, the Bluejays are at least as likely to win the World Series this year as that team is."

This is obviously a weaker claim than (39):

---

14 The validity of this inference is in fact built into the axioms of comparative possibility by Kratzer's predecessor Lewis (1973) and by Halpern (1997), whose discussion is inspired by Lewis. However, the damaging consequences of the Disjunctive Inference for Kratzer's theory do not seem to have been noted in the literature until they were independently discovered by me (Lassiter 2010) and by Seth Yalcin (Yalcin 2010b). As Halpern notes, this property is shared by several other representations of uncertainty, e.g. possibility logic and fuzzy logic. The problem noted here is shared by these frameworks, as far their usefulness for modeling natural language is concerned.

(39)   "The Bluejays are at least as likely to win as they are not to win."

But on Kratzer's theory, (38) plus a few simple facts about baseball actually ENTAILS (39). This is unacceptable: the weaker statement should not entail the stronger.

Let $\mathbf{T} = \{\mathbf{team}_1, ..., \mathbf{team}_{29}\}$ be the other 29 Major League Baseball teams. (40) is a reasonable rendition of the truth-conditions of (38):

(40) $\forall x \in \mathbf{T}$: It is at least as likely that the Bluejays win as it is that x wins.

Let p be the proposition *The Bluejays win*, and let q_n be the proposition *Team$_n$ wins*. (40) is equivalent to (41):

(41) $(p \geqslant^s q_1) \wedge (p \geqslant^s q_2) \wedge ... \wedge (p \geqslant^s q_{29})$

Feeding (41) into schema (37) repeatedly, we see that (42) follows as well.

(42) $p \geqslant^s (q_1 \vee q_2 \vee ... \vee q_{29})$

Since one of the thirty teams must win, the only way that the Bluejays fail to win is if someone else does — that is, *The Bluejays do not win* is true iff $(q_1 \vee q_2 \vee ... \vee q_{29})$ is true. So we can rewrite (42) as (43):

(43)   *The Bluejays win* is at least as likely as *The Bluejays do not win*.

Putting this all together: (38) and the rules of baseball entail (43), which is equivalent to (39). But (39) is clearly a much STRONGER claim that (38), not an entailment.

In general, in Kratzer's theory a disjunction ϕ, no matter how large, can never be as likely as another proposition ψ unless one of the disjuncts of ϕ is itself as likely as ψ. Is this a plausible prediction about valid inferences from sentences involving *likely*? The answer seems to be "no". Somehow, a disjunction of lower-ranked possibilities must be able to "gang up" to overpower a higher-ranked possibility.

Both of the problems discussed in this section involve the fact that Kratzer's theory lacks QUANTITATIVE information about likelihood. This suggests that we need a semantics for epistemic modals which employs quantitative measures. Numerical probability, I will show, provides precisely what we need.

4.4   Probabilistic semantics for gradable epistemic modals

We are looking for a semantics for epistemic modals which has the following three properties: it associates GEMs with a fully closed scale, renders quantitative statements like (36) meaningful, and does not validate the Disjunctive Inference (37). A well-understood logical system which fits the bill — the probability calculus.

(44)   A *Probability Space* is a pair $\langle W, \mu \rangle$, where W is a set of possible worlds and $\mu : \wp(W) \to [0, 1]$ is a function from subsets of W to real numbers between 0

and 1 which satisfies the following conditions:[15]

a. $\mu(W) = 1$.

b. If $P \cap Q = \varnothing$, then $\mu(P \cup Q) = \mu(P) + \mu(Q)$.

Re-naming μ as *prob*, (rough) truth-conditions for the GEMs follow immediately.

(45)   a. ϕ *is possible* is true iff $prob(\phi) \neq 0$.

b. ϕ *is likely/probable* is true iff $prob(\phi) > \mathbf{s}_{poss}$.

c. ϕ *is certain* is true iff $prob(\phi) = 1$.

And, of course, truth-conditions for comparative possibility sentences:

(46) ϕ *is more likely than* ψ is true iff $prob(\phi) > prob(\psi)$.

(45) and (46) are readily derived compositionally from a slightly modified version of Kennedy & McNally (2005)'s semantics for gradable adjectives; see Lassiter 2010 for details. (These conclusions were foreshadowed in unpublished work by Seth Yalcin mentioned in §3 above, and to some extent also in Horn 1989: §5.)

There are several important connections between the definitions in (44) and the theory of gradability. First, the definitions as stated entail the existence of a bottom element 0 in addition to the top element 1. Just as in K&M's theory, though, the choice of the numbers 0 and 1 as bottom and top elements is arbitrary.

In addition, note that axiom (44b) is not a special stipulation needed for probability measures. Rather, it is a general feature of EXTENSIVE MEASUREMENT (Krantz et al. 1971), and holds, for example, of measurements of height, length, and fullness:

(47)   a. If you place two (disjoint) boards of length m and n end-to-end, the total length is $m + n$.

b. If you pour a cup that is $n\%$ full into a (different) cup of the same volume that is $m\%$ full, the result is a cup that is $(n + m)\%$ full.

So the definitions in (44) are not special features designed to solve the problems we observed with Kratzer's theory: they follow directly from a very general theory of measurement, on the assumption that the scale relevant to epistemic modals is fully closed. In other words, the use of probability to give scales for GEMs is a very natural product of the general theory of measurement and the fact that the scale of possibility has a maximum ($prob(W)$) and a minimum ($prob(\varnothing)$).

Our first desideratum is clearly satisfied: probability provides a fully closed scale. On the second issue, it is simple to give meaningful truth-conditions for the sentence-types in (36) (and possible to derive them compositionally as well):

---

15 Note that (44) assumes that W is finite. The definitions are a bit more involved when W is infinite, but this complication is not important here.

(48)　　a. ϕ *is twice as likely as* ψ is true iff $prob(\phi) = 2 \times prob(\psi)$.

　　　　b. *It is half certain that* ϕ is true iff $prob(\phi) = 0.5 \times 1 = 0.5$.

　　　　c. *It is 95% certain that* ϕ is true iff $prob(\phi) = 0.95$.

Finally, the unwanted Disjunctive Inference comes out invalid. Counter-model: let $prob(\phi) = .3$, $prob(\psi) = .2$, $prob(\chi) = .2$, and $prob(\psi \wedge \chi) = 0$. Then $prob(\psi \vee \chi) = .3 + .2 = .5$, which is greater than $prob(\phi)$.

5   Conclusion

A good theory of gradable epistemic modals should have at least these three features: it should be compatible with a good theory of the compositional semantics of gradable adjectives; it should not render valid any clearly invalid inferences; and it should predict all of the clearly valid inferences.

The standard Kratzerian theory of modality, despite its other merits, is difficult to integrate with a theory of gradability, makes demonstrably incorrect predictions about inferences involving disjunction and equatives (37), and fails to yield meaningful truth-conditions for relatively unremarkable sentence-types like those in (36). In contrast, a theory based on probability satisfies these desiderata: it is readily compatible with K&M's theory of gradability, it gives denotations to the sentence-types in (36), and it does not validate (37). By all current indications, a probabilistic semantics is the best option for the gradable epistemic modals considered here.

Many important questions remain unsettled, though. For instance, other gradable epistemic modals such as *clear*, *evident*, and *apparent* have not been addressed, and it is unclear to what extent probability is implicated in their semantics (Barker 2009). A second interesting question is whether there are other natural language expressions whose semantics make reference to probability: for some proposals, see van Rooij 1999, Levinson 2003, and Cohen 1999.

Another interesting question that we have not addressed is whether epistemic modal auxiliaries also have a probabilistic semantics. On the one hand, parsimony considerations suggest that option is worth considering. On the other hand, Kratzer's theory has the not inconsiderable virtue of explaining the varied uses of modal auxiliaries like *should*, *must*, *may*, and *ought* without needing to posit massive lexical ambiguity. One option is simply to allow Kratzerian semantics for modal auxiliaries and probabilistic semantics for GEMs to exist side-by-side, or even attempt to derive probability measures from comparative possibility (see Kratzer 2010 for relevant discussion). A second option is to embrace probabilistic semantics for epistemic modal auxiliaries as well, and to explain their apparent non-gradability as due to syntactic restrictions. If we take this approach, we have the further burden of providing a theory that can match Kratzer's in explaining the semantic variability

of modal auxiliaries without positing ambiguity. This would mean devising a scale-based theory of deontic and circumstantial modals (see Lassiter 2010).

Finally, there are many questions of detail about the probabilistic analysis, including what interpretation of probability is appropriate, exactly what types of information — and whose — are relevant, whether probability statements have ordinary truth-conditions (Yalcin 2010a), and whether even probability is a rich enough representation format (Halpern 2003). However these questions are resolved, though, it seems clear that numerical probability — or something with its core properties — plays a crucial role in the semantics of epistemic modals.

References

Barker, Chris. 2009. Clarity and the grammar of skepticism. *Mind and Language* 24(3). 253–273. doi:10.1111/j.1468-0017.2009.01362.x.

Cohen, Ariel. 1999. Generics, frequency adverbs and probability. *Linguistics and Philosophy* 22(3). 221–253. doi:10.1023/A:1005497727784.

Halpern, Joseph. 1997. Defining relative likelihood in partially-ordered preferential structures. *Journal of Artificial Intelligence Research* 7. 1–24. doi:10.1613/jair.391.

Halpern, Joseph. 2003. *Reasoning about Uncertainty*. MIT Press.

Heim, Irene. 2001. Degree operators and scope. In Fery & Sternefeld (eds.), *Audiatur vox sapientiae: A festschrift for Arnim von Stechow*, Berlin: Akademie Verlag.

Horn, Laurence. 1989. *A Natural History of Negation*. University of Chicago Press.

Kennedy, Chris. 2001. Polar opposition and the ontology of 'degrees'. *Linguistics and Philosophy* 24(1). 33–70. doi:10.1023/A:1005668525906.

Kennedy, Chris. 2007. Vagueness and grammar: The semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy* 30(1). 1–45. doi:10.1007/s10988-006-9008-0.

Kennedy, Chris & Louise McNally. 2005. Scale structure, degree modification, and the semantics of gradable predicates. *Language* 81(2). 345–381. doi:10.1353/lan.2005.0071.

Krantz, David H., R. Duncan Luce, Patrick Suppes & Amos Tversky. 1971. *Foundations of Measurement*. Academic Press.

Kratzer, Angelika. 1981. The notional category of modality. In Hans-Jürgen Eikmeyer & Hannes Rieser (eds.), *Words, worlds, and contexts: New approaches in word semantics*, 38–74. de Gruyter.

Kratzer, Angelika. 1991. Modality. In Arnim von Stechow & Dieter Wunderlich (eds.), *Semantics: An international handbook of contemporary research*, de Gruyter.

Kratzer, Angelika. 2010. *Modality and Conditionals*. Oxford University Press (To Appear).

Lassiter, Daniel. 2010. *Measurement and Modality*: New York University dissertation. (In preparation).

Levinson, Dmitry. 2003. Probabilistic model-theoretic semantics for *want*. In Robert B. Young & Yuping Zhou (eds.), *Semantics and Linguistic Theory (SALT) XIII*, 222–239. CLC Publications.

Lewis, David. 1973. *Counterfactuals*. Harvard University Press.

Portner, Paul. 2009. *Modality*. Oxford University Press.

Potts, Christopher. 2008. Interpretive Economy and Schelling points. Ms., University of Massachusetts at Amherst .

van Rooij, Robert. 1999. Some analyses of pro-attitudes. In H. de Swart (ed.), *Logic, game theory, and social choice*, Tilburg University Press.

Rotstein, Carmen & Yoad Winter. 2004. Total adjectives vs. partial adjectives: Scale structure and higher-order modifiers. *Natural Language Semantics* 12(3). 259–288. doi:10.1023/B:NALS.0000034517.56898.9a.

Sapir, Edward. 1944. Grading: A study in semantics. *Philosophy of Science* 11(2). 93–116. doi:10.1086/286828.

Sassoon, Galit. 2010. The degree functions of negative adjectives. *Natural Language Semantics* 18(2). 141–181. doi:10.1007/s11050-009-9052-8.

Schwarzschild, Roger & Karina Wilkinson. 2002. Quantifiers in comparatives: A semantics of degree based on intervals. *Natural Language Semantics* 10(1). 1–41. doi:10.1023/A:1015545424775.

Yalcin, Seth. 2007. Epistemic modals. *Mind* 116(464). 983–1026. doi:10.1093/mind/fzm983.

Yalcin, Seth. 2010a. Nonfactualism about epistemic modality. In A. Egan & B. Weatherson (eds.), *Epistemic modality*, Oxford University Press.

Yalcin, Seth. 2010b. Probability Operators. *Philosophy Compass* 916–937.

Daniel Lassiter
Department of Linguistics
New York University
10 Washington Place
New York, New York 10003
lassiter@nyu.edu