POSITIVE AND CONDITIONAL SEMANTICS FOR GRADABLE MODALS*

PETER KLECHA
University of Chicago

1 Introduction

This paper presents a semantics for gradable modal adjectives (GMAs), focusing particularly on *likely*. Lassiter (2010, 2011) argues that the adjectives *possible*, *likely*, and *certain* have a common, gradable, probability-based semantics, on the basis of entailment relations between these terms. This claim contradicts Kennedy's (2007) Interpretive Economy which predicts that adjectives with the same scales should have the same positive interpretations. I argue that these adjectives do not provide a counterexample to Interpretive Economy, and give a semantics for them which preserves the entailment relations discussed by Lassiter, while also preserving the general theory of modality (Kratzer 1981, 1986); to this end I show that one important aspect of Kratzer's analysis, the analysis of conditionals as domain restrictors (Kratzer, 1986) can be preserved for gradable modals.

1.1 Interpretive economy

Kennedy & McNally (2005) argue scales come in four types: **upper-closed**, **lower-closed**, **fully closed**, and **fully open**.

(1) a. FULLY CLOSED SCALE: $[0,1]$
b. UPPER-CLOSED SCALE: $(0,1]$
c. LOWER-CLOSED SCALE: $[0,1)$
d. FULLY OPEN SCALE: $(0,1)$

According to Kennedy's theory, there is a single (null) positive morpheme, *pos*, which combines with the gradable adjective, which denotes a simple measure function, and derives a property. The *pos* morpheme picks out a standard on the scale associated with the measure function it combines with, and returns the property of having a greater degree of the relevant gradable property than the standard. The standard is denoted by *pos* as being the lowest degree on the relevant scale which *stands out* on that scale.

*Thanks to Rebekah Baglini, Max Bane, M. Ryan Bochnak, Adrian Brasoveanu, Lucas Champollion, Itamar Francez, Tim Grinsell, Chris Kennedy, Greg Kobele, Dan Lassiter, Lisa Matthewson, Ora Matushansky, Yaron McNabb, Galit Sassoon, Stephanie Solt, Malte Willer, and my SuB 16 audience.

It is this notion of *stands out*, in combination with Interpretive Economy, which derives the apparent variation in positive meanings.

(2) a. The cup is full. (=totally full) (MAXIMAL)
b. The boy is tall. (=relatively tall) (RELATIVE)
c. The nail is bent. (=not totally straight) (MINIMAL)

Kennedy's Interpretive Economy says that if the *stands out* notion can be derived from only the conventional aspects of the meanings of the expressions in play, then the value for *pos* will be determined according to those conventional aspects. Only in the absence of such conventional touchstones will interlocutors resort to context to determine the value of the positive standard.

Closed scales have such conventional touchstones which interlocutors can reliably converge upon to determine what value on the scale stands out most. Thus if an adjective is associated with a scale with a maximum, its positive meaning is maximal; if an adjective is lower-closed and not upper-closed its positive meaning is minimal; and if an adjective is fully open it is relative, i.e., its positive meaning is determined by context.

Since scale type entirely determines positive meaning, this analysis predicts that no two adjectives should ever share a scale but have different positive readings. Lassiter (2011), however, argues that the trio *possible/certain/likely* falsify this prediction. Lassiter argues that all three have a common scale: the probability scale, which is a fully closed scale. However he argues that rather than all three having maximal positive readings, *possible* is minimal and *likely* is relative.

(3) a. That scenario is possible. (prob > 0) (MINIMAL)
b. That scenario is likely. (prob $\geq s_c$) (RELATIVE)
c. That scenario is certain. (prob = 1) (MAXIMAL)

This paper argues against Lassiter's claims about the scale structural properties of these three expressions. In Section 2, I argue against the treatment of *possible* as a gradable adjective, arguing that its apparently gradable behavior can be explained without appeal to a measure function analysis. In Section 3, I argue that *certain* and *likely* do not share a common scale, as *certain* should be instead associated with a confidence scale. In Section 4, I argue that *likely* is not lexically associated with a closed scale, although it may underlyingly relate to a closed scale like the probability scale. In Section 5 I will explore the consequences of the claims about *likely*'s scale structure for theories of modality and conditionals, and conclude in Section 6.

1.2 Diagnostics

Before presenting my arguments, I summarize the diagnostics I will apply in determining the status of the various expressions under discussion.

Gradability is determined by assessing the compatibility of the expression with degree modifiers. Gradable expressions crucially are of a type $\langle \alpha, d \rangle$, whereas non-gradable predicative adjectives are of type $\langle \alpha, t \rangle$. Degree modifiers (DMs) are of the right type to combine only with gradable adjectives. Thus, gradable expressions should be robustly acceptable with a wide range of DMs, and non-gradable adjectives should not.

(4) a. The ball is {bigger/very big/too big/so big/etc...}
b. The linguist is {*deader/*very dead/*too dead/*so dead/etc...}

Big is judged to be good with a wide range of DMs, *dead* is judged to be bad with the same DMs; so we may conclude that *big* is gradable and *dead* is not. To supplement this judgment data, a corpus search was conducted to determine the relative frequencies of each DM with a sample of adjectives; see Appendix A for details.

Degree modification with *dead* and other non-gradable adjectives is not totally unattested, however, (see Appendix A) and attestations of NGAs with degree modification cannot all be explained as speech errors. I propose that such attestations are cases of coercion, where NGAs behave like GAs so long as some kind of scale can be readily associated with their conventional, non-gradable interpretations. Consider, for example, (5).

(5) The patient cannot be resuscitated; he is too dead.

Here it can be imagined that the speaker means something like "it has been too long since he died", or "his body is too damaged". Conversely, the speaker might mention a patient who is not too dead to resuscitate, i.e., the patient has no heartbeat but could be defibrillated.

Scale type, the second major semantic property of interest, is determined by compatibility with scale-specific degree modifiers. The semantics of these modifiers depends crucially on a particular scale structure; so compatibility with them is diagnostic of that structure.

Upper-closed-scale adjectives are acceptable with maximizing degree modifiers (MDMs), which pick out the maximum on the adjectival scale. MDMs are therefore infelicitous if the scale has no maximum. The MDMs I examine are *totally*, *completely*, *perfectly*, and *absolutely*.

(6) a. The cup is **totally** {full/dry/#big/#dirty}.
b. The cup is **completely** {full/dry/#big/#dirty}.
c. The cup is **perfectly** {full/dry/#big/#dirty}.
d. The cup is **absolutely** {full/dry/#big/#dirty}.

Crucially, compatibility must be assessed with respect to the *maximizing reading* of the modifiers. For example, *totally* and *completely* can have partitive readings; e.g., (6-b) can mean something like "all of the nail is bent". These readings must be excluded from this diagnostic.

Next, fully-closed-scale adjectives are acceptable with proportional degree modifiers (PDMs). PDMs take proportions of scales, which require the scale to be fully closed. The PDMs I examine are *n percent*, fractions like *three-quarters* and *half* (Bochnak, 2010), and modifiers like *all-the-way* and *partially*. As with maximizing modifiers, partitive uses must be excluded.

(7) a. The cup is **{10/50/100}%** {full/#dry/*big/*dirty}.
b. The cup is **{three quarters/half/one third}** {full/#dry/*big/*dirty}.
c. The cup is **{all/most/half}-(of)-(the)-way** {full/#dry/*big/*dirty}.
d. The cup is **mostly/partially** {full/#dry/*big/*dirty}.

Finally, lower-closed scale adjectives (which are not also upper-closed) are compatible with minimizing degree modifiers (MinDMs) *slightly* and *a little*. MinDMs crucially can also take an excessive reading with any kind of adjective (Bylinina 2011, Solt 2011) which must be excluded.

(8) a. #The cup is **slightly** {#full/#dry/#big/dirty}.
b. #The cup is **a little** {#full/#dry/#big/dirty}.

Open-scale adjectives do not comport with any of these diagnostics. Thus, these diagnostics show that *full* and *dry* are upper-closed adjectives, *full* is fully-closed, *dirty* is lower-closed, and *big* is fully open.

2 Against a common probabilistic core

In this section I argue against the claim that *possible*, *certain*, and *likely* share a common scale, the probability scale. I will do this by arguing that *possible* is not a gradable adjective at all, and that *certain*, while gradable, does not map to the probability scale.

2.1 *Possible*

Lassiter claims *possible* to be a minimum-scale gradable adjective on the data in (9), among others.

(9) a. It is slightly possible that the Jets will win.
b. It is possible the Jets will win, but it could be more possible.

Lassiter (2010:203) notes: “Many speakers accept the comparative *more possible*, though some express discomfort, preferring *more likely* [...] 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.”

However, this finding conflicts with judgment data I have collected, and the corpus data Lassiter mentions could not be reproduced; cf. the attestation frequencies in Appendix A.

(10) It is {*more/*very/*so/*too/$^{\text{OK}}$quite/*rather/etc...} possible that the ball is in his left hand.

Note that the attestation values presented in Appendix A are much closer to the average for non-gradable adjectives than for gradable ones; for example, only 0.18% of uses of *possible* were in the comparative, compared to 9% for all gradable adjectives searched and 18% for *likely*. On the other hand, non-gradable adjectives appeared with the comparative morpheme 0.15% of the time, bearing a striking resemblance to *possible*.[1]

Given this finding, a defender of the proposal that *possible* and its antonym are gradable bears an extraordinary explanatory burden. And while Lassiter demonstrates that *more possible* is indeed attested, this does not contradict the finding: every other non-gradable adjective, with the exception of *existing*, was attested at least once with many degree modifiers. However, as discussed above, this can be accounted for as cases of coercion.

One exception to the generalization, interestingly, is *quite*. Acceptability judgments with *quite possible* show that it is indeed acceptable, and this is confirmed by attestation data. However, a few things in the corpus findings are worth pointing out: first, the average non-gradable adjective appears with *quite* 0.08% of the time; the average gradable adjective appears with *quite* 0.15% of the time, around twice as frequently. This is the narrowest margin of any degree modifier, which suggests that *quite* is not a very good predictor of gradability. Moreover, *possible* appears with *quite* 1% of the time, which is not only much more than any other non-gradable adjective examined,

[1] Lassiter also points to *impossible*, the antonym of *possible* as a gradable expression. Corpus findings for *impossible* are similar to *possible*, however (see Appendix).

(i) *It is {more/very/too/etc...} impossible that the ball is in his left hand.

but in fact it is much more than any gradable adjective examined. This suggests *quite possible* is a relatively idiosyncratic case. Note that, expectedly, *quite impossible* is also more frequent than the average for gradable adjectives, though still much less frequent than *quite possible*.

Further discussion of *quite possible* is beyond the scope of this paper, but I will point out that an analysis which treats *quite* in *quite possible* as an idiosyncratic modifier not derivationally related to the more general DM *quite* is much simpler than positing *possible* as gradable, but idiosyncratically not combining with 13 of the 14 major English DMs.

This homophonous, idiosyncratic *quite* might have the effect of an overt ordering source, narrowing the modal base of a traditional quantificational modal, giving rise to a stronger modal claim. This would explain the apparent similarity to intensification.

(11) $[\![\textbf{possible}]\!] = \lambda p \,.\, \lambda m_{\text{s,st}} \,.\, \lambda w \,.\, \exists v \in m(w) \,.\, p(v)$

(12) $[\![\textbf{quite}_2]\!] = \lambda M_{\langle \text{st}, \langle \langle \text{s,st} \rangle, \langle \text{s,t} \rangle \rangle \rangle} \,.\, \lambda p \,.\, \lambda m_{\text{s,st}} \,.\, \lambda w \,.\, M(\textsc{best}(p)(m))(w)$

I leave aside a number of issues, such as: what the precise semantics for the ordering is (see Kratzer (2012), Lassiter (2011)), whether unmodified *possible* has its own ordering source apart from *quite*, and what role context plays in all of this; these go beyond the scope of this paper; for a more detailed discussion see Klecha (forthcoming). The analysis as given here is sufficient to demonstrate that a non-gradable semantics can plausibly be given for *possible*. One outstanding issue is the entailment relations between *possible* and *likely*; I discuss these in Section 4.

2.2 *Certain*

Lassiter (2010) argues that *certain* is a GMA relying on a probability scale, and having a maximum positive reading, consistent with Kennedy (2007). He observes the following data.

(13) a. It is completely certain that the Jets will win.
b. #It is certain that the Jets will win, but it could be more certain.

This comports with my own diagnostics for gradability and upper-closed scales, although diagnostics for fully-closed scales are less clear. First, note robust acceptability with degree modifiers[2], acceptability with MDMs, and unacceptability with MinDMs, consistent with Lassiter's claims.

(14) a. It is {more/very/too/etc...} certain that the ball is in his left hand
b. It is {totally/completely/perfectly/absolutely} certain that Herman Cain will lose.
c. #It is {slightly/a little} certain that Herman Cain will lose.

Interestingly, acceptability judgments for proportional modifiers are far less certain.

(15) a. ?It is {10/50/100}% certain that Herman Cain will lose.
b. ?It is {three quarters/half/one third} certain that Herman Cain will lose.
c. ?It is {all/most/half}-(of)-(the)-way certain that Herman Cain will lose.

[2] A corpus search for *certain* was not conducted due to confusability with the homophonous adjective *certain*, as in *a certain book*. This *certain* appears to be highly frequent and thus would muddle any searches which only restrict by grammatical category. However, the acceptability of *certain* with degree modifiers is not a subject of present controversy.

d. ?It is mostly certain that Herman Cain will lose.

Consider particularly that as the values for *n%* decrease, so does acceptability. It is not even clear what (16-c) is intended to mean. [3] It is unclear to me what the correct analysis of this phenomenon is; but it does at least throw into question *certain*'s status as a fully-closed scale adjective. Compare with *full*, which is clearly acceptable with any value for *n%*.

(16) a. It is 95% certain.
b. ?It is 60% certain.
c. *It is 30% certain.

(17) a. It's 95% full.
b. It's 60% full.
c. It's 30% full.

Moreover, Lassiter's claim that *likely* and *certain* share the same (probability) scale seems problematic. Consider the following.

(18) a. Obama's victory couldn't be less certain $\rightarrow$ Obama's defeat couldn't be less certain
b. Paul's victory couldn't be less likely $\rightarrow$ Paul's defeat couldn't be more likely

If *couldn't be less* is a kind of minimizing modifier[4] which picks out the lowest point on a scale, this seems to suggest the following: *zero certainty* of p entails *zero certainty* of $\neg p$, while *zero likelihood* of p entails *maximal likelihood* of $\neg p$. This would strongly suggest that *certainty* and *likelihood* are not synonymous; rather, *certain* and *likely* denote different scales.

I propose that *certain* employs a *confidence* scale, whereas *likely* employs something closer to probability scale (more on that below). I propose that the confidence scale orders propositions which some agent already believes or considers plausible or likely and ranks them by that agent's confidence in the evidence for each proposition. An alternative, pointed out to me by Lucas Champollion (p.c.) and developed in some detail by Lassiter (2011) in response to the arguments made here, is to analyze *certain* in terms of the information-theoretic notion of entropy[5].

Deciding among these alternatives is beyond the scope of this paper; see Klecha (forthcoming). In any case it remains clear that *likely* and *certain* are not associated with the same scale.

3 *Likely*

Since *possible* is not gradable and *certain* uses a different scale, there is no longer a counterexample from GMAs to the prediction that there should never be scalemate gradable adjectives which have differing positive readings. However, there is still the claim by Lassiter that *likely* is a fully-closed-scale adjective but also a relative adjective, apparently contrary to Kennedy (2007). However, degree-modification evidence shows that *likely* is an open-scale adjective.

[3] A limited corpus search seems to confirm this; a search was conducted for *percent certain*, which yielded 44 hits. Because the total number of hits for the sense of *certain* we are concerned with is unknown, it is hard to evaluate the meaning of this. However, of the 44 hits, all but one were for values between 90 and 100, and the last remainder was for *80 percent certain*. In other words, there were no hits for *sixty percent certain*, *forty percent certain*, etc.

[4] Lassiter (2011) rightly points out that *couldn't be less* is not a perfect device for getting at the minimum value on a scale, based on sentences like *He couldn't be less friendly*, since it is not obvious that *friendly* has a lower bound. However, the intuition that (17) gets at is sound, namely that *certainty* and *likelihood* do not not perfectly correlate.

[5] Lassiter also considers using the information-theoretic notion of *surprisal*, but rejects this possibility owing to the fact that it would derive an upper-closed scale, rather than a fully-closed scale as desired. However, given the data above, it seems plausible that *certain* is in fact only upper-closed, in which case Lassiter's proposal regarding surprisal may become more attractive.

3.1 Degree modification

First, it is widely agreed that *likely* is gradable, and is not a lower-closed adjective.

(19) a. It is {more/less/very/too/so...} likely that Santorum will lose.
b. #It is {slightly/a little} likely that Mitt Romney will lose.

Consider *likely* with MDMs and PDMs, which are diagnostic of upper- and fully-closed scales, respectively.

(20) #It is {totally/completely/perfectly/absolutely} likely that Herman Cain will lose.

(21) a. ?It is {10/50/100}% likely that Herman Cain will lose.
b. #It is {three quarters/half/one third} likely that Herman Cain will lose.
c. #It is {all/most/half}-(of)-(the)-way likely that Herman Cain will lose.
d. #It is mostly likely that Herman Cain will lose.

First, *likely* is bad with maximizing degree modifiers, contrary to what is expected if it is a fully closed scale adjective. Lassiter (2010) actually points this out as well, but argues that maximizing degree modifiers do not diagnose scales with maxima (more on this claim below).

Furthermore, *likely* is also clearly bad with most proportional modifiers. One potential exception is *n%*, which Lassiter argues to be acceptable with *likely* and in fact to be crucially diagnostic of a fully-closed scale. Intuition judgments vary, but there are indeed attestations of *n% likely*.[6] However, even if *n%* is acceptable with *likely*, it is unclear why other PDMs don't work; e.g., why shouldn't (22-a-c) and (23-a-c) have the same acceptability?

(22) a. ?It's 50% likely.
b. *It's half likely.
c. *It's halfway likely.

(23) a. ?It's 75% likely.
b. *It's three-quarters likely.
c. *It's three-quarters-of-the-way likely.

I argue that in the case of *likely*, *n%* is a measure phrase rather than a proportional modifier. In other words, *n%* denotes some number of a unit of likelihood, rather than denoting a proportion. This means that *n%* is not a diagnostic for a closed scale[7]. This explanation might also be applicable to *certain*, which displayed a puzzling pattern with regard to *n%* modification, although it is still unclear how exactly to explain that phenomenon.

Lassiter (2010) attempts to explain contrasts like (22-a-b), by arguing that DMs select not only for scale type but for positive meaning – thus *n%* is the only true diagnostic of closed scales because it selects only for a closed scale adjective; the other proportional degree modifiers, and apparently all the maximizing ones, select for fully-closed-scale adjectives *whose positive meanings are maximal*. Since *likely* idiosyncratically has a relative positive meaning (all positive meanings are

[6] A corpus search could not conclusively decide this matter. A search for "percent likely" yielded 2 hits, out of 8,402 total hits for *likely*; however, a search for "percent full" yielded only 12 hits out of 11,760 total hits for *full*. This means that *percent full* is about 5 times more common than *percent likely*; however, a statistical analysis has not been run on these terms and the overall numbers are quite small, so gauging significance is difficult. In the absence of conclusive evidence against it, I will continue on the assumption that *n% likely* is acceptable and must be explained. However, if it can be shown that *n% likely* is not acceptable, this would only simplify my case.

[7] The measure phrase *n%* is taken to be homophonous with the PDM *n%* which has the usual semantics.

determined idiosyncratically on Lassiter's take, which follows from Kennedy & McNally (2005)) it cannot combine with any other maximizing or degree modifiers outside of *n%*.

This raises some questions, however. First, why is *n%* the only such modifier? It would strengthen Lassiter's claim if even one other such modifier could be found. Moreover, *likely* seems to be the only adjective with a fully-closed scale *and* a relative positive meaning. Lassiter's theory is unmotivated on this count and therefore is disprefered to the one presented here.

3.2 Bounds and maxima and minima

By the diagnostics given, *likely* appears to have an open-scale, which has the virtue of aligning with Kennedy (2007) in light of it being a relative adjective. However, there is the apparent drawback that the proposed semantics for *likely* seems to not relate to the intuitive, albeit mathematically sophisticated notion of probability which involves a closed scale; indeed, Lassiter seems to take this intuitive notion of probability as being the primary motivator for pursuing a closed-scale analysis.

But ultimately the "intuitive scale" associated with an adjective does not always align with its lexical scale, nor should we expect it to. An adjective may have as its basic scale something fairly intuitive, but then build arbitrary (or non-arbitrary) presuppositions or constraints which alter the scale (structure). Or an adjective's meaning may simply be unintuitive.

Particularly, consider *short* and *inexpensive*, which Lassiter (2010) cites as further counterexamples to Kennedy (2007). Lassiter argues that *short*'s scale has a maximum value, since there is obviously (intuitively) a maximum to shortness (likewise a minimum for tallness; in other words there is a minimum height, namely 0). Since these terms are relative, not minimal and maximal respectively, these may seem like counterexamples to Kennedy (2007). Likewise, *inexpensive* has an obvious maximum, namely a cost of 0, but is not a maximum adjective.

However, just because these intuitive scales are closed, we should not conclude that the lexical scales are. First of all, maximizing modifiers do not work with *short* or *inexpensive*.

(24) a. #The boy is {totally/completely/perfectly/absolutely} short.
b. #The car is {totally/completely/perfectly/absolutely} inexpensive.

Certainly these cannot mean "the boy has a height of 0" or "the car is free", respectively. This is troubling for an account which says these adjectives' scales have maxima. Second, elements that occupy the apparent endpoints of these scales are not obviously admissible with these expressions. In the case of *short*, it seems very clear that elements with zero height are not ordered; after all, these items do not take up space.

(25) #That blade of grass is very short, but taller than dignity (since dignity does not have physical extent).

While *short* may be upper-bounded, it does not have a maximum element, which is the crucial factor[8]. This requires a modification to Kennedy & McNally's (2005) typology of scales; in addition to the usual categories discussed in Section 1.1, scales may be upper-bounded but lack a maximal element; lower-bounded but lack a minimal element, or both.

(26) a. FULLY BOUNDED: (0,1)

[8]Similar things can be said about any gradable adjective relating to spatial extent: *long*, *wide*, *small*, etc.

b. UPPER-BOUNDED/LOWER-OPEN: $(-\infty,1)$
c. UPPER-BOUNDED/LOWER-CLOSED: $[0,1)$
d. LOWER-BOUNDED/UPPER-OPEN: $(0,\infty)$
e. LOWER-BOUNDED/UPPER-CLOSED: $(0,1]$

As for *inexpensive*, Dan Lassiter (p.c.) points out that compared to (25), it doesn't seem quite as bad to use it to refer to something which has no cost.

(27) ?Breathing air is inexpensive; in fact, it's free.

I argue, however, that *inexpensive* really does exclude free objects from its ordering, but examples like (27) are really cases of coercion – i.e., scales with no inherent maxima or minima can be coerced to include such end points in the ordering to the extent that it is intuitive to do so. This can also explain the contrast between *Dignity is short* and *Breathing air is inexpensive*; it is far less of a stretch to imagine *breathing air* as being subject to cost than to imagine dignity as having height, thus, coercion is easier in the first case than the second. Non-gradable adjectives are subject to the same ease-of-coercion cline; *prime* is very hard to coerce, *pregnant* relatively easy.

(28) a. ?Carissa is very pregnant (in fact, she's almost due.)
b. ??This patient is not very dead (he could easily be resuscitated.)
c.???6 is less prime than 8. (...?)

So if *likely* relies upon a fully-closed scale with no minimum or maximum (a scale from 0 to 1 exclusive, or (0,1)) we can preserve *likely* as using something like a probability scale but whose extremes are excluded from the ordering (see Section 4.4). This predicts the oddity of (29).

(29) ??It is more likely that Obama will be reelected than that 2+2=5.

As with *inexpensive*, (29) is good to the extent that the expression can be coerced to include maximum or minimum values in the ordering.

4 Likelihood and modality

The biggest question surrounding any scalar modal semantics is: how, if at all, does it relate to the non-scalar quantificational modal semantics established in the work of Kratzer? Lassiter (2011) argues for a strong rejection of Kratzer's theory on the basis that a scalar modal semantics cannot easily be related to a quantificational one, and that even if it could, a uniform theory is preferred, going as far as to even argue for a scalar approach to modal auxiliaries like *must* and *might*. I argue for a more conservative approach which maintains Kratzer's theory.

4.1 Core Kratzer

The general theory of modality outlined by Kratzer (1981, 1986) says the following: modals are essentially relations between a modal base[9] and a prejacent proposition; the modal base is a highly contextually variable element (Relative Modality); the modal base can be subject to scalar

[9]In strict Kratzerian terminology, the modal base is a contextually derived set of propositions, which are intersected to yield a set of worlds which form the quantificational domain, modulo the effect of ordering sources and *if*-clauses, for a quantificational modal. Here I will mostly use "modal base" in a looser sense, simply meaning the domain

restrictions (Ordering); and conditionals are derived by restricting the modal base (Conditionals as Restrictors). These core proposals have formed the backbone of significant research in modality.

Additional claims were made by Kratzer as well which have since been shown inadequate. One is Kratzer's apparent claim that all modal expressions could be reduced to simple quantificational expressions involving $\forall$ or $\exists$; another is her denotation for *likely* which does not allow for a compositional account of degree modification. Neither of these are core claims of Kratzer's, but they are consequences of the analysis presented in Kratzer (1981) which have been carried forward in much research on modality. However, as discussed in Yalcin (2010), Lassiter (2011), and Kratzer (2012), these two proposals are inadequate for dealing with clearly gradable modal expressions like *likely*, which must denote measure functions.

4.2 The denotation

Nevertheless, the existence of gradable modals is not a problem for this core Kratzerian approach to modality. I argue for the denotation for *likely* in (30-a). This denotation, and the denotations of other gradable modals which may follow its schema, is parallel to the basic schema for categorical modals in (30-b). In (30-a), $lhood$ is an additive[10] measure function taking a modal base $m(w)$ and a prejacent p and returning the likelihood of p in $m(w)$.

(30) a. $[\![\textbf{likely}]\!] = \lambda p \,.\, \lambda m_{\text{s,st}} \,.\, \lambda w \,.\, lhood(m(w))(p)$
b. $[\![\text{categorical modal}]\!] = \lambda p \,.\, \lambda m_{\text{s,st}} \,.\, \lambda w \,.\, Q(m(w))(p)$

Both kinds of expressions take a prejacent p, an accessibility relation m, and an evaluation world w, where $m(w)$ is a modal base; gradable modals return a degree, categorical modals a truth value. Categorical modals relate the modal base and prejacent by a generalized quantifier; gradable modals, on the other hand, denote a measure function, which maps modal base-prejacent pairs to a degree on a scale. Gradable modals may differ in terms of what exactly the ordering is, and in terms of what modal base-prejacent pairs are included in the ordering.

4.3 Entailment relations

One major argument Lassiter (2011) uses in trying to argue for a unified-scale approach to *possible*, *likely*, and *certain* is the entailment patterns they show.

(31) a. p is certain $\rightarrow$ p is likely
b. p is likely $\rightarrow$ p is possible
c. p is not possible $\rightarrow$ p is not likely
d. p is not likely $\rightarrow$ p is not certain

If these three adjectives make use of the same scale, and their positive forms indicate different, ordered points on that common scale, these entailment relations fall out naturally. However, these entailment patterns can also be derived quite naturally without such an assumption.

of a modal, the set of worlds which represent the background information against which a prejacent proposition is evaluated; for a quantificational modal, this means that "modal base" = "quantificational domain".

[10]See Yalcin (2010) and Lassiter (2011) for discussion of and arguments for Additivity, which they also impose as a constraint on *likely*'s ordering. I define Additivity as follows, diverging slightly from Yalcin and Lassiter in including the modal base in the formulation:

(i) *Additivity*: $p \cap q \cap m(w) = \emptyset \rightarrow lhood(m(w))(p \vee q) = lhood(m(w))(p) + lhood(m(w))(q)$

First, (31-a) can be derived if the scales for *likely* and *certain* relate in some way. I proposed briefly that *certain* only orders propositions taken by some individual to be likely in the first place. Lassiter's (2011) entropy proposal discussed above builds a complex measure function out of the more basic probability measure used by *likely* – also ensuring a natural relation between the two.

The bigger issue is the entailment relation between *likely* and *possible* – we cannot rely on a common scalar core to derive these entailments. In fact, Lassiter uses other such entailments to argue for uniform scalarity across the modal domain, including modal auxiliaries like *must* and *might*. He argues that while these terms are not of the right type to be gradable for arbitrary syntactic reasons, their core semantics is indeed scalar.

However, as we have seen, *possible* is not gradable. This fact cannot be blamed on its syntactic category, which undercuts Lassiter's argument that modal auxiliaries are only non-gradable for syntactic reasons. (I do not discuss Lassiter's approach to auxiliaries any more here, but see Klecha (forthcoming).) Moreover, a common scalar core is clearly not required to derive entailment relations (e.g., universals entail existentials without reference to degrees), although the measure function for *likely* needs to have the right properties. In fact, given the denotation above, it does.

First, since *lhood* is additive, if $q \cap m(w) \subseteq p \cap m(w)$ then $lhood(m(w))(p) > lhood(m(w))(q)$; in other words, if p is entailed by q in the modal base $m(w)$, then its likelihood is greater than q's (given $m(w)$). Put yet another way, bigger sets of worlds receive higher *lhood* values. It follows directly from Additivity, then, that if $p \cap m(w) = \emptyset$, then $lhood(m(w))(p)$ has the lowest value that *lhood* could assign, say, 0. In other words, if p is not possible in the classical sense (there are no worlds in the modal base in which p is true), then p has no likelihood.

Second, as discussed by Lassiter (2011), *likely* is a relative adjective, so the truth of ϕ *is likely* depends on the contextually determined threshold or standard. However, regardless of the choice of standard, exceeding the standard would entail a non-zero likelihood, which in turn entails truth in at least one world. These facts, then, derive the entailments in (31-b-c).

4.4 Conditionals

Any theory of gradable modals must account for their behavior in the conditionals, like in (32).

(32) If Drew picks black, he is likely to win.

On my analysis, this can be accounted for within the Restrictor Theory of Conditionals. Crucial to this fact is the presence of the modal base in my analysis, which is targeted by *if*. Assuming the modal base is present (but silent) in the tree, the denotation for *if* in (33) suffices.

(33) $[\![\textbf{if}]\!] = \lambda q \,.\, \lambda m \,.\, \lambda w \,.\, m(w) \cap q$

if as defined above combines with its complement, then with the modal-base-denoting variable in the syntax, intersecting the modal base with the antecedent clause[11], giving a conditional reading.

Yalcin (2010) argues that Kratzer's (1986) Restrictor Theory must be amended slightly to account for *likely*. He argues that the evaluation function, for all expressions, is indexed to a modal domain and to a probability measure. The function of *if* is to restrict this modal domain in

[11] This simplifies the picture somewhat, by excluding discussion of the ordering source and its potential effect on which antecedent worlds are included in the modal domain.

the usual Kratzerian way, but to also "conditionalize" the probability measure on the truth of the antecedent.

The first issue with this approach is that it applies the solution to *conditional probability* to *if* itself; this is a case of overfitting, since this conditional probability measure is vacuous for all expressions except *likely* and perhaps a few others which involve probability, as Yalcin (2010) himself points out. Second, this conditionalized probability measure is not needed given the restriction of the modal base. Restriction of the modal base is sufficient if the probability measure has the right properties.

In fact, the probability measure proposed by Yalcin (2007), amended to derive probability relative to a modal base rather than all worlds, is just such a measure. Yalcin's proposal is that *likely* is associated with a partitioning function Π which, given a prejacent p, partitions W such that in every partition ι, p is either true throughout ι or false throughout ι; a probability measure Pr then assigns each partition a probability such that probabilities of the partitions sum to 1. The likelihood of p is then the sum of the probabilities of all the p-partitions. I adapt this proposal so that Π partitions the modal base (X in (34-a)) rather than the set of all words, W.

(34) a. $\Pi_{\mathrm{p,X}}$ partitions X such that $\forall\iota \in \Pi_{\mathrm{p,X}}.[\iota \subseteq p] \vee [\iota \cap p = \emptyset]$ and $\sum_{\iota \in \Pi_{\mathrm{p,X}}} Pr(\iota) = 1$

b. $lhood(m(w))(p) = \sum_{\iota \in \Pi_{\mathrm{p,m(w)}} \cap \mathrm{Pow(p)}} Pr(\iota)$

The result is that restricting the modal base via the conditional means that Pr will assign probabilities to the partitions of the restricted modal base in such a way that the sum of these partitions' probabilities is 1. So if q is true in all $m(w) \cap p$-worlds, i.e., if *if p then q in w* is true, then $lhood(m(w) \cap p)(q)$ = 1, the desired result for conditional probability. The probability measure itself need not be conditionalized; rather, probability is always relative to a domain.

The contrast with Yalcin (2010) is a subtle one, but it is important to show a) that conditionals work the same for gradable and nongradable modals, rather than stipulating an additional effect for *likely* and b) that the modal base/modal domain is crucial and its role in the modal semantics must be made clear. In doing so it can be shown that gradable and nongradable modals, despite their differences, share the basic properties of modals established in Kratzer's work.

Finally, returning to the discussion of scale structure in the previous section, note that if we also constrain Π such that there must always be at least one partition in which the prejacent is true and one in which the prejacent is false, we derive the constraint against propositions ever being mapped to what are in principle the maximum and minimum values on the scale, 1 and 0, respectively, giving us the scale structure consistent with the distribution of *likely* discussed above. Thus likelihood is associated in some sense with a probability scale [0,1], but its lexical properties prohibit maximum or minimum values, deriving a bounded scale (0,1).

5 Conclusion

This paper shows that the scalar properties of gradable modal adjectives do not present counterexamples to Kennedy's (2007) theory of Interpretive Economy. Moreover, semantics can be given for these expressions which allow for some modal expressions to be gradable and others non-gradable without sacrificing any empirical ground, particularly regarding entailment relations between gradable and non-gradable properties, and while maintaing a uniform approach to conditionals.

References

Bochnak, M. Ryan. 2010. Two sources of scalarity in the verb phrase. *In:* Boban Arsenijević, Berit Gehrke, & Marín, Rafael (eds), *Subatomic Semantics of Event Predicates* (to appear). Dordrecht: Springer.

Bylinina, Lisa. 2011. “Functional standards" and relative/absolute distinction. This volume.

Davies, Mark. 2008. *The Corpus of Contemporary American English: 425 Million Words, 1990-Present*. http://corpus.byu.edu/coca.

Kennedy, Chris, & McNally, Louise. 2005. Scale Structure and Semantic Typology of Gradable Predicates. *Language*, **81**(2), 345–81.

Kennedy, Christopher. 2007. Vagueness and grammar: the semantics of relative and absolute gradable adjectives. *Linguistics and Philosophy*, **30**(1), 1–45.

Klecha, Peter. forthcoming. *Scales and Modality*. Ph.D. thesis, University of Chicago.

Kratzer, Angelika. 1981. The notional category of modality. *Pages 38–74 of:* Eikmeyer, H.-J., & Rieser, H. (eds), *Words, Worlds, and Context*. de Gruyter.

Kratzer, Angelika. 1986. Conditionals. *Proceedings of CLS*, **22**(2), 1–15.

Kratzer, Angelika. 2012. *Modality and Conditionals*. Oxford University Press.

Lassiter, Daniel. 2010. Gradable epistemic modals, probability, and scale structure. *SALT* **20**.

Lassiter, Daniel. 2011. *Measurement and Modality*. Ph.D. thesis, New York University.

Solt, Stephanie. 2011. Comparison to arbitrary standards. This volume.

Stanley, Jason. 2000. Context and Logical Form. *Linguistics and Philosophy*, **23**(4), 391–434.

Yalcin, Seth. 2007. Epistemic Modals. *Mind*.

Yalcin, Seth. 2010. Probability Operators. *Philosophy Compass*, **5**(11), 916–37.

A Corpus Search Results for Adjectives

A series of corpus searches were conducted using the Corpus of Contemporary American English (COCA) (Davies, 2008). The corpus data in COCA is sorted by source-type, i.e., spoken, fiction, etc; all searches are done only on spoken data. COCA is annotated for grammatical category; so all searches exclude homophonous expressions of different categories. Below are the results for the searches conducted. Each figure in the table represents the percentage of uses of the adjective which were coupled with the degree modifier. For example, when assessing the ability of *dog* to combine with the definite determiner *the*, a percentage is given, indicating the percent of instances of *dog* which are preceded by *the*, in this case 0.290. Averages are given for gradable adjectives (AVG:GA) and non-gradable adjectives (AVG:NGA); both averages exclude expressions under discussion, namely *likely*, *possible*, and *impossible*.

Table 1: Adjectives with common degree modifiers.

	TOTAL	more/-er	very	most/-est	how	too	so	enough
big	23961	0.0717	0.0561	0.0392	0.0277	0.0268	0.0232	0.0161
small	21857	0.1469	0.0657	0.0205	0.0030	0.0116	0.0086	0.0021
tall	1959	0.1210	0.0470	0.0755	0.0311	0.0133	0.0112	0.0061
long	29559	0.0779	0.0425	0.0248	0.0059	0.0113	0.0184	0.0008
short	10099	0.0616	0.0696	0.0129	0.0007	0.0123	0.0051	0.0003
important	36900	0.0607	0.1796	0.1250	0.0219	0.0037	0.0352	0.0023
likely	8402	0.1790	0.0390	0.0865	0.0061	0.0012	0.0010	0.0001
impossible	24113	0.0018	0.0002	0.0008	0.0021	0.0005	0.0020	0.0000
possible	16250	0.0018	0.0095	0.0001	0.0003	0.0000	0.0002	0.0000
dead	10768	0.0027	0.0007	0.0003	0.0000	0.0000	0.0003	0.0000
American	56702	0.0030	0.0008	0.0011	0.0010	0.0002	0.0004	0.0001
right	51657	0.0006	0.0006	0.0001	0.0001	0.0001	0.0010	0.0000
pregnant	3567	0.0008	0.0048	0.0006	0.0014	0.0000	0.0008	0.0000
alive	6495	0.0022	0.0022	0.0015	0.0005	0.0002	0.0018	0.0002
existing	1198	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AVG:GA		0.0900	0.0767	0.0497	0.0151	0.0132	0.0169	0.0046
AVG:NGA		0.0015	0.0015	0.0006	0.0005	0.0001	0.0007	0.0000

Table 2: Adjectives with less common degree modifiers.

	TOTAL	as	pretty	that	this	quite	rather	less
big	23961	0.0160	0.0130	0.0041	0.0009	0.0005	0.0004	0.0003
small	21857	0.0033	0.0039	0.0009	0.0001	0.0021	0.0017	0.0000
tall	1959	0.0158	0.0056	0.0031	0.0015	0.0031	0.0005	0.0015
long	29559	0.0015	0.0016	0.0034	0.0001	0.0005	0.0010	0.0000
short	10099	0.0030	0.0025	0.0003	0.0000	0.0011	0.0016	0.0002
important	36900	0.0146	0.0021	0.0025	0.0000	0.0016	0.0003	0.0040
likely	8402	0.0171	0.0011	0.0001	0.0000	0.0050	0.0000	0.0551
impossible	24113	0.0026	0.0005	0.0011	0.0000	0.0027	0.0001	0.0003
possible	16250	0.0020	0.0001	0.0016	0.0000	0.0100	0.0000	0.0002
dead	10768	0.0020	0.0002	0.0000	0.0000	0.0002	0.0001	0.0000
American	56702	0.0014	0.0000	0.0000	0.0000	0.0001	0.0000	0.0001
right	51657	0.0001	0.0000	0.0000	0.0000	0.0045	0.0000	0.0001
pregnant	3567	0.0003	0.0000	0.0000	0.0000	0.0000	0.0003	0.0000
alive	6495	0.0009	0.0000	0.0003	0.0000	0.0000	0.0000	0.0000
existing	1198	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
AVG:GA		0.0090	0.0048	0.0024	0.0005	0.0015	0.0009	0.0010
AVG:NGA		0.0008	0.0000	0.0001	0.0000	0.0008	0.0001	0.0000