This paper examines the modal expressions *possible*, *likely*, *probable*, and *certain* (henceforth GEMs). Tests for scale structure suggest that they denote functions from propositions to degrees on an upper- and lower-bounded scale. Support for taking this scale to be the range of numerical probabilities [0,1] comes from examples with disjunction and existential quantification. I present evidence that Kratzer (1981)'s semantics of comparative possibility makes incorrect predictions for these cases, while numerical probability makes correct predictions.

As Portner (2009) notes, a semantics for gradable modals should be compatible with a general theory of gradable expressions. Following Kennedy (2007), gradable adjectives denote functions from individuals to degrees on a scale: a triple \( (D, \prec, \delta) \), where \( D \) is a set of degrees, \( \prec \) a total ordering of \( D \), and \( \delta \) the dimension of the adjective (e.g., “closure”). Degree modifiers impose conditions on the degree denoted by a measure function, as in (1).

\[
\begin{align*}
\text{(1)} & \quad \text{a. } [\text{closed}] = \lambda x. \lambda d. [x’s degree of closure = d] \\
& \quad \text{b. } [\text{completely closed}] = \lambda x. \text{id}. [x’s degree of closure = d \land d = \max(D_{\text{closed}})]
\end{align*}
\]

Portner considers an analysis of GEMs as probability operators, i.e. as denoting functions from propositions to degrees on the scale \( ([0,1], \prec, \text{likelihood}) \). However, Portner is skeptical that GEMs denote on the same scale, because their degree modifiers are not uniform. But adjectives associated with the same scale may accept different modifiers (Kennedy 2007):

\[
\text{(2)} \quad \text{The room is completely full/?#occupied.}
\]

In Kennedy’s terms, this difference is due to the fact that *full* requires that an object possess a maximal degree of fullness, while *occupied* merely requires that an object possess a non-zero degree. I will argue that the distribution of degree modifiers with GEMs has a similar explanation: the differences are not because they denote on different scales, but because they denote different points on the same scale. I analyze each GEM in turn, showing that the distribution of degree modifiers is what we expect if they denote probability operators corresponding to Kennedy’s minimum-, maximum-, and relative-standard adjectives.

In terms of Kennedy’s typology, *certain* is a maximum-standard adjective like *full*: it requires its argument to have a (near-)maximal degree of likelihood, and is associated with the upper portion of an upper-bounded scale. This is shown by the fact that *certain* behaves like *full* on various tests. For example, both can be modified by *completely*, but not *slightly*.

\[
\begin{align*}
\text{(3)} & \quad \text{a. The room is completely/#slightly full.} \\
& \quad \text{b. It is completely/#slightly certain that Thunderbolt will win the race.}
\end{align*}
\]

*Possible* is a minimum-standard adjective, like *occupied* or *bent*. Minimum-standard adjectives apply to objects that have a non-zero degree of the property in question, and are associated with the lower portion of a lower-bounded scale. Kennedy claims that, if an adjective can be modified by *slightly*, it falls in this class.

\[
\begin{align*}
\text{(4)} & \quad \text{a. Do slightly bent spokes matter? (google)} \\
& \quad \text{b. It’s slightly possible that an asteroid could trigger a nuclear war. (google)}
\end{align*}
\]
Likely and probable fall among the the **relative-standard** adjectives such as tall. Unmodified relative adjectives have a “greater than contextual standard” semantics (via a silent morpheme pos); and, according to Kennedy, they are odd with both completely and slightly.

(5) a. Mary is #completely/#slightly tall.
b. It is #completely/#slightly likely that Thunderbolt will win.

The facts we have seen show that GEMs are associated with a scale which has an upper and a lower bound. Why should we take this scale to be the scale of numerical probabilities? One reason is that, for a certain class of examples, probability makes correct predictions where its primary competitor, comparative possibility à la Kratzer (1981), does not.

Imagine you are watching a horse race. Horse A is in the lead, but B,C,D, and E are close behind. You might be inclined to agree with (6a), but be doubtful about (6b)-(6c):

(6) a. A is more likely to win than B, and A is more likely to win than C, and ... than E.
b. It is more likely that A will win than it is that B or C or D or E will win.
c. It is more likely that A will win than it is that another horse will win.

However, Kratzer’s semantics predicts that no rational person should be able to make this judgment: in fact, all the statements in (6) are logically equivalent for her.

(7) \( p \) is more possible than \( q \), \( p \gtrsim q \) (relative to a modal base \( f \) and an ordering source \( g \)) iff:

- a. \( \forall u \in \bigcap f(w) : (u \in q) \rightarrow \exists v \in \bigcap f(w) : v \gtrsim g(w) \land v \in p \).
- b. \( \exists u \in \bigcap f(w) : (u \in p) \land \neg \exists v \in \bigcap f(w) : v \in q \land v \gtrsim g(w) \land u \).

(Kratzer 1981:48)

According to (7), a proposition is exactly as likely as the most likely world(s) it contains (this is the effect of the existential quantification in (7b)). Thus, \( p \) is more likely than \( q \) iff the top-ranked world in \( p \) outranks the top-ranked world in \( q \). Assuming that \( [p \lor q] = [p] \cup [q] \), it follows from (7) that a disjunction is exactly as likely as its most likely disjunct, and thus that (6a) and (6b) are equivalent. And since (6b) = (6c) if there are no other horses, Kratzer’s semantics predicts, against intuition, that all the sentences in (6) are equivalent.

In contrast, in a probability-based semantics (6a) is not equivalent to (6b) or (6c):

(8) a. \( [(6a)] = [\text{prob}(A \text{ wins}) > \text{prob}(B \text{ wins})] \land ... \land [\text{prob}(A \text{ wins}) > \text{prob}(E \text{ wins})] \)
b. \( [(6b,c)] = 1 \text{ iff } \text{prob}(A \text{ wins}) > \text{prob}([B \text{ wins}) \lor (C \text{ wins}) \lor (D \text{ wins}) \lor (E \text{ wins}) \lor (F \text{ wins})] \)
   \[ = 1 \text{ iff } \text{prob}(A \text{ wins}) > \Sigma_{x \neq A} \text{prob}(x \text{ wins}) \]

(8a) is true and (8b) false, e.g., if \( \text{prob}(A \text{ wins}) = .4 \) and \( \text{prob}(x \text{ wins}) = \frac{1-4}{4} = .15 \) for \( x \in \{B,C,D,E\} \). Intuitively, this situation is possible; the probability-based approach can model it, but Kratzer’s semantics of comparative possibility cannot.

Facts about degree modification show that a probability-based approach is possible; facts about disjunctions and quantified statements with likely show that probability fares better than its main competitor. Thus, it appears, probability yields the right semantics for GEMs.

(9) a. \( [p \text{ is pos likely/probable}] = 1 \text{ iff } \text{prob}(p) > s_{\text{likely}} \) (the contextual standard for likely).
b. \( [p \text{ is possible}] = 1 \text{ iff } \text{prob}(p) > \min([0,1]) \) (i.e., if \( \text{prob}(p) > 0 \)).
c. \( [p \text{ is certain}] = 1 \text{ iff } \text{prob}(p) = \max([0,1]) \) (i.e., if \( \text{prob}(p) \simeq 1 \)).