Plan:

1. Kratzer’s relativity likelihood analysis of probability operators.
2. Hamblin’s possibility-theoretic analysis of probability operators.
3. A probability-theoretic analysis.

0 Preliminaries

0.1 Syntactic brief

I’m interested in the semantics of what I’ll call probability operators:

- **likely**: mostly performs as adjective with an infinitival complement in raising constructions (1), but has adverbial uses (2), can occur predicatively with a finite clausal complement and an expletive subject (3), can modify nouns within DPs (4), and occurs in comparative form with all the usual degree morphology (5).

  1. Adj: Bob is (very, most) likely to be/will be/have been home.
  2. Adv:
     i. Bob (very, most) likely is home/was home/will be home.
     ii. Most likely Bob is home.
  3. Expletive: It’s (very, most) likely that Bob is home/was home/will be home.
  4. Attributive: Lou Dobbs is a likely presidential candidate.
  5. Overt comparative:
     i. Bob is (at least) as likely to be at the party as Steve.
     ii. Bob is (much) more/less likely to be at the party than Steve.
     iii. Bob’s being at the party is more likely than/is as likely as Steve’s being at the party.

- **probable**: an adjective which dislikes raising environments (6) and adverbial uses (7), preferring expletive constructions (8), attributive uses (9), and comparative forms (10).

  6. Adj: * Bob is probable to be/will be/have been home.
  7. Adv: * Bob probable is home/was home/will be home.
  8. Expletive: It’s (very) probable that Bob is home/was home/will be home.
  9. Attributive: Recession is a probable outcome of these fiscal policies.
  10. Overt comparative:
      Bob’s being at the party is more probable than/is as probable as Steve’s being at the party.

- **probably**: a sentential adverb that can occur preverbally in a simple monoclausal sentence:

  11. (i) Bob probably left.  (ii) Probably Bob left.
0.2 Common semantic assumptions

- Probability operators express operations on propositions.
- \( \text{probably } \phi \iff \text{it is probable that } \phi \iff \text{it is likely that } \phi \)
- \( \phi \text{ is at least as likely as } \psi \iff \phi \text{ is at least as probable as } \psi \)
- The comparative form—in particular, \( \text{is at least as likely/probable as} \), abbreviate it \( \succeq \)—is the conceptually basic notion. In terms of it, we can express the meaning of the other comparatives:

\[
\begin{align*}
\phi \text{ is as probable/likely as } \psi & \iff \phi \succeq \psi \text{ and } \psi \succeq \phi \\
\phi \text{ is more probable/likely than } (\succ) \psi & \iff \phi \succeq \psi \text{ and } \psi \nsucceq \phi
\end{align*}
\]

and we can analyze the superficially one-place \( \text{probably/it is likely that} \) as a covert comparative:

\[\text{Probably } \phi \iff \phi \succ \neg \phi\]

It is fairly standard in the gradable adjectives literature to take the comparative form as expressing the theoretically basic notion (cf. Kennedy [2007]).

1 First semantics: relative, qualitative likelihood (Kratzer 1991)

1.1 The theory

- Kratzer [1981, 1991] groups probability operators semantically with modals. Her motivation (I take it) is the logical connection of \( \text{probably} \) to epistemic modals:

\[
\begin{align*}
\text{V1. Must to probably.} \\
& \text{It must be that } \phi \\
& \text{Probably } \phi
\end{align*}
\]

\[
\begin{align*}
\text{V2. Probably to might.} \\
& \text{Probably } \phi \\
& \text{It might be that } \phi
\end{align*}
\]

We might appeal also to parallels in their distributional properties. For instance, like epistemic modals, probability operators prefer to scope above quantifiers (von Fintel and Iatridou [2003]):

(12) ? Everyone probably lost the lottery.

So Kratzer's account of probability operators is situated within her grand plan for modality generally. On this approach, modal expressions (\( \text{might, can, ought, should, must, possibly, etc.} \)) are analyzed in the metalanguage as effecting quantification over possible worlds relative to two contextually supplied parameters:

a. Modal base. One parameter—the modal base—delivers the domain of modal quantification. The fact that some modals (\( \text{may, ought, etc.} \)) allow for a multiplicity of interpretations (deontic, epistemic, nomological, etc.) is explained by appeal to permissible variability in this parameter. In the special case of probability operators, the relevant modal base will supply a set of epistemically accessible worlds, a set of worlds left open by the evidence available in the world at which the clause is evaluated for truth.\(^1\)

\(^1\) Alas, often it is not easy to say whose evidence is at issue in unembedded uses of epistemic modal sentences. See MacFarlane [2006], Yalcin [2007b], von Fintel and Gillies [2008]. I will completely ignore this issue.
b. **Ordering source.** The second parameter—the *ordering source*—imposes a preorder $\geq$ over the worlds supplied by the modal base. In the case of probability operators, the worlds are to be intuitively understood as preordered by relative likelihood.\(^2\)

- To state the semantics for the comparative *is at least as probable/likely as* (using $\geq$ now for the relation this comparative expresses) we use the preorder (reflexive transitive order) $\geq$ over worlds supplied by the ordering source to define a preorder on propositions. This might be done in various ways. Kratzer’s approach is a very natural one:\(^3\)

$$ p \geq q \iff \forall w \in q : \exists w' \in p : w' \geq w $$

(Understand our quantifiers as restricted already by the relevant epistemic modal base.) A proposition $p$ is *as likely as* $q$ (with respect to a preordered set) just in case every $q$-world in the set is matched by a $p$-world in the set at least as high in the preorder. The strict comparative *is more likely than* is then defined in terms of $\geq$ as follows:

$$ p \succ q \iff (p \geq q) \land \neg (q \geq p) $$

and the one-place operator *probably* is defined in terms of $\succ$ to mean ‘is more likely than not’.

- So:

1. context supplies a set of *epistemically accessible worlds* and a set of propositions to serve as the *ordering source*;
2. the ordering source induces a *preorder on possible worlds* in the modal base;
3. this preorder on worlds induces the *preorder on propositions* expressed by *is at least as probable/likely as*;
4. the rest of the probability operators are defined in terms of *is as likely as*.

### 1.2 Strengths of the account

- Kratzer’s account looks pretty baroque—mainly because it does not appeal to quantities or magnitudes of probability, only a qualitative ordering over worlds. With these very sparse resources, however, it covers an impressive range of intuitively valid inference patterns. Understanding consequence in terms of preservation of truth at a world, given a fixed context, it is trivial to prove that Kratzer’s semantics validates:

\[ \begin{align*}
V3. & \text{ Probably to not probably not.} \\
& \text{probably } \phi \\
& \text{ } \neg \text{ probably } \neg \phi \\
V4. & \text{ Probably-distribution over conjunction.} \\
& \text{probably } (\phi \land \psi) \\
& \text{probably } \phi \land \text{ probably } \psi \\
V5. & \text{ Chancy disjunction introduction.} \\
& \text{probably } \phi \\
& \text{probably } (\phi \lor \psi)
\end{align*} \]

---

\(^2\)On Kratzer’s technical development, this preorder is not taken as primitive. Rather the ordering source delivers, at the world of evaluation, a set of $O$ of propositions, and this set is then used to induce the preorder on worlds as follows: $w \geq_O w' \iff \{ p \in O : w \in p \} \supseteq \{ p \in O : w' \in p \}$. $O$ is meant to be a set of propositions that is ‘normally’ or ‘stereotypically’ true in the relevant situation. (This idea is not much elaborated in Kratzer [1991], but see Kratzer [1989].)

\(^3\)Technical features of this qualitative model of probability are independently discussed by Halpern [2003].
V6. Minimality

\[ \phi \text{ is at least as likely as } \bot \]

V7. Maximality

\[ \top \text{ is at least as likely as } \phi \]

Together with her semantics for epistemic modals (Kratzer [1991]), Kratzer secures Must to probably (V1) and Probably to might (V2) above. Together with her semantics for indicative conditionals (Kratzer [1986]), she validates:

V8. Chancy Modus Ponens.

\[
\begin{array}{c}
\text{if } \phi \text{ then } \psi \\
\text{probably } \phi \\
\text{probably } \psi
\end{array}
\]


\[
\begin{array}{c}
\text{if } \phi \text{ then } \psi \\
\neg \text{ probably } \psi \\
\neg \text{ probably } \phi
\end{array}
\]


\[
\begin{array}{c}
\text{if } \phi, \text{ then } \psi \\
\psi \text{ is at least as likely as } \phi
\end{array}
\]

(Conditional to comparative might seem surprising. Consider denying the conclusion but accepting the premise: “Sally’s being at the party is more likely than Steve’s being there; but if Sally is at the party, then Steve is.” That is jarring. Plausibly, ‘\( \phi \) is more likely than \( \psi \)’ entails ‘it might be that \( (\phi \land \neg \psi) \)’, which (in a fixed context) seems intuitively incompatible with ‘if \( \phi \), then \( \psi \)’.)

1.3 Two eyebrow-raising inference patterns, and how Kratzer’s account settles them

1.3.1 Conjunctivitis

• First a pattern I’ll call, after Kyburg [1970], Conjunctivitis:

E1. Conjunctivitis.

\[
\begin{array}{c}
\text{probably } \phi \\
\text{probably } \psi \\
\text{probably } (\phi \land \psi)
\end{array}
\]

According to Kratzer’s account (Kratzer [1991]), Conjunctivitis is invalid. (The failure of this inference pattern would make it clear that probably is not amenable to direct analysis as the \( \Box \)-operator of a normal modal logic,\(^4\) an interesting consequence.) Is this the right call?

• Very often, this inference sounds fine, as observed by Hamblin [1959]. Indeed, it is not easy to come up with completely natural-sounding counterexamples. But it is easy to imagine the sort of cases which \textit{would} be counterexamples to this inference pattern, if the correct semantics for the probably-operator represented it as expressing that its propositional complement has a (probability axiom-respecting) probability above some threshold.

Consider a fair twelve-sided die, with the sides numbered one to twelve. The die is rolled. How did it come up? Reflecting on the fact that there are more sides numbered below 9 than not, you may be inclined to agree that:

\[^4\text{Since in a normal modal logic, it is a theorem that } (\Box \phi \land \Box \psi) \supset \Box (\phi \land \psi).\]
(P1) Probably, a number below 9 came up.
and, by parity, that:

(P2) Probably, a number above 4 came up.
yet if so, you are probably not also tempted to agree that:

(C) Probably, a number above 4 and below 9 came up.

- It should be remarked, however, that many feel a tension when asked to assume (P2) immediately after (P1). This tension is suggestive of the phenomenon of modal subordination (Roberts [1989]), wherein a modal expression occurring later in a discourse is interpreted as ‘anaphoric’ on an earlier modal, so that the clause it operates on is understood as contributing information to an already-described possibility. If the probably in (P2) is interpreted in this way, the total discourse (P1)-(P2) is just equivalent to (C), and the corresponding inference is trivially valid—explaining why the pattern often strikes us as fine.

Upshot: hard to probe the validity of this inference pattern “directly”; let theory decide.

1.3.2 Apparent failure of Modus Tollens

- Behold:

E2. Apparent Failure of Modus Tollens.

\[
\text{if } \phi, \text{ then probably } \psi \\
\neg \text{ probably } \psi \\
\neg \phi
\]

If Sally came to the party, Steve probably came; but it is not likely that Steve came. From these assumptions, does it follow that Sally didn’t come to the party?

No; from such wishy-washy premises, nothing so cut and dry follows. Intuitively, the two premises are both compatible with Sally’s having gone to the party. If anything, what follows from the premises is merely that it’s not likely that Sally went to the party.

- “Apparent?” Whether this is a bonafide failure of Modus Tollens depends on whether the probably-operator in the first premise is to be rightly understood as a component of the consequent of the conditional at the relevant level of analysis. Ways to argue that things are otherwise:

1. **Wide-scoping.** Accord to the first way, probably is really taking scope over the whole conditional in the first premise above; i.e., its structure is PROBABLY(\(\phi \rightarrow \psi\)).

Comment: this is unsatisfying, for the question remains: what rules out the possibility of the scopal order \(\rightarrow\), PROBABLY? That question could be skipped if we could detect a difference between these two allegedly logically possible scopal orders, and then simply declare that intuitively, the scopal order in our first premise is really PROBABLY, \(\rightarrow\). But evidently we cannot. If there are two scopal orders theoretically possible, it seems that they are equivalent:

\[[\phi \rightarrow \text{probably } \psi] \iff [\text{probably } (\phi \rightarrow \psi)]\]

Without a semantic difference between the orders, it is hard to see how to make the case that PROBABLY is taking wide scope. More natural, it seems, to suppose that the actual scopal order is the superficial one—the one problematic for Modus Tollens.

2. **Restrictor analysis of if (Kratzer).** More promising would be to follow Kratzer [1986], who dispenses with the idea that there is a scope-taking conditional operator to begin with. Rather, she models if-clauses semantically as restrictors on a (sometimes tacit) modal operator. On her approach, the syntactic analysis of our first premise would be something more like:
PROBABLY$_\phi(\psi)$

—making it clear that probably is not a part of the consequent after all. Modus Tollens would be safe on this analysis.

1.4 Problems

The range of inference patterns Kratzer is able to validate given her meager assumptions is remarkable. But alas: there are valid inference patterns Kratzer’s account fails to validate, and there are patterns the account validates which are clearly invalid.

1. **Valid patterns not validated.** Consider:

V11. Positive Form Transfer

\[ \psi \text{ is at least as likely as } \phi \]
\[ \text{Probably } \phi \]
\[ \text{Probably } \psi \]

Example: “Steve is at least as likely as Bob to be at the party, and Bob is probably at the party; so Steve is probably at the party.” This form is clearly valid. On Kratzer’s account this form is equivalent to the (likewise valid):

V12. Complement Transfer

\[ \psi \text{ is at least as likely as } \phi \]
\[ \phi \text{ is at least as likely as } \neg \phi \]
\[ \psi \text{ is at least as likely as } \neg \psi \]

Example: “Steve is at least as likely as Bob to be at the party, and Bob is at least as likely as not to be at the party; so Steve is at least as likely as not to be at the party.” This form is more convoluted, but is nevertheless very plausible.

Neither of these inference patterns are valid on Kratzer’s semantics.

5

2. **Invalid patterns validated.** The **union property** is had by all models of uncertainty underwriting this pattern:

U1. Union property pattern.

\[ \phi \text{ is at least as likely as } \psi \]
\[ \phi \text{ is at least as likely as } \chi \]
\[ \phi \text{ is at least as likely as } (\psi \lor \chi) \]

In fact, Kratzer’s semantics validates this pattern. (Halpern [2003], though he doesn’t discuss Kratzer or the semantics of probability operators, observes that the relatively likelihood model she uses satisfies this property.) This is an egregious result:

---

5Consider an epistemic modal base totally ordered (chained) by some ordering source. Assume that the chain ascends into \( p \) infinitely (i.e., that there is a world \( w \) in the order such that for every \( w' \geq w \), \( p(w') = 1 \)), and that as it ascends, it alternates between \( q \) and \( \neg q \) infinitely. For Kratzer, this is a case where \( p \) is probable (indeed, epistemically necessary), and \( q \) is as likely as \( p \), but \( q \) is not probable.

The counterexample does not depend on the assumption of totality, which is made for simplicity, but it does rely on the failure of a limit assumption for the relevant ordering. Kratzer rejects the limit assumption; but would a limit assumption help us avoid the problem? No, for we can adapt the counterexample. Imagine again a total order, and consider the set \( M \) of worlds which are maximal according to the order. Let \( p \) be true throughout \( M \), but \( q \) be true only in a proper subset of \( M \). Here again \( p \) will be probable (indeed, epistemically necessary), and \( q \) is as likely as \( p \), but \( q \) will not be probable (since for every \( q \)-world, there is a \( \neg q \) world at least as high up in the order).

What if we make the limit assumption and the uniqueness assumption (the assumption that there are no ties in likelihood between worlds)? Kratzer could then avoid the counterexample, but at a prohibitive cost: this would collapse the semantics of probably into her semantics for epistemic must, and it would leave it dark how to capture gradations in probability (as when we say “\( p \) is unlikely, but more likely than \( q \)”). The move also has little conceptual motivation: why shouldn’t ties in likelihood between worlds be allowed?
The coin’s landing heads is at least as likely as the coin’s landing tails.
The coin’s landing heads is at least as likely as the coin’s landing heads.
So, the coin’s landing heads is at least as likely as the coin’s landing heads or tails.

This seems like obviously invalid reasoning.

Reply: “You misread the conclusion. The conclusion actually employs wide scope ‘or’,
so in fact it is fine. And good luck trying to force a narrow scope reading.”

Reply to reply: The reply assumes that a narrow scope ‘or’ reading of the
conclusion is impossible, and this seems implausible. Anyway, the underlying
problem can be set out without having to find a wide-scope ‘or’ test. A special
case of (U1), at least under classical assumptions, is:

\[ \text{U2. Bizarro consequence of the union property pattern.} \]
\[ \phi \text{ is as likely as } \phi \]
\[ \phi \text{ is as likely as } \neg \phi \]
\[ \phi \text{ is at least as likely as } \top \]

This is already insane, but we can sharpen the problem; the inference U2 +
Maximality (that \( \top \) is at least as likely as any proposition) + the fact that ‘\( \phi \)
is as likely as \( \phi \)’ is a tautology gives us:

\[ \text{U3. Bizarro consequence of the union property pattern, 2.} \]
\[ \phi \text{ is as likely as } \neg \phi \]
\[ \phi \text{ is at least as likely as } \psi \]

So, if the coin’s landing heads is as likely as its landing tails (not heads), it is,
therefore, as likely as any proposition you please! Reductio. Kratzer’s account
doesn’t work.

2 Second semantics: possibility spaces (Hamblin 1959)

2.1 The theory

• It is interesting that Kratzer does not even consider the question of using a conventional prob-
ability model in the semantics of probability operators. In contrast, Hamblin [1959] opens with
that question:

Metrical probability-theory is well-established, scientifically important and, in essen-
tials, beyond logical reproof. But when, for example, we say “It’s probably going to
rain”, or “I shall probably be in the library this afternoon”, are we, even vaguely, using
the metrical probability concept? (234)

He goes on to build a case for a negative answer to this question, developing a semantic proposal
inspired by the non-additive representations of uncertainty then being explored by the economist
G.L.S. Shackle. Shackle argued that traditional probability theory was too idealized to problems
arising in “the real environment of human life”, where we find that “there is, instead of clear-cut
simplicity of the games of chance, a fog of ignorance and confusion arising not from remediable
shortcomings of human organization... but arising from the nature of things...” ([Shackle, 1953,
99]).

\[ \text{Shackle was in turn heavily influenced by Keynes, who also argued that in many real-life cases of interest to economists} \]
\[ \text{(e.g., the expected time until the obsolescence of an invention, the prospect of war, the rate of interest in twenty years,} \]
\[ \text{etc.) there was just “no scientific basis on which to form any calculable probability whatsoever”.} \]
• In the spirit of this kind of view, Hamblin defined a new kind of measure for representing uncertainty, one less restrictive in certain respects than that of ordinary probability. The class of measures he defined are nowadays called **possibility measures**.

Suppose for simplicity that $W$ is a finite set of worlds. A **possibility measure** $\text{Poss}$ is any function assigning to each subset of $W$ (i.e., the propositions constructible from $W$) a number in $[0,1]$ satisfying the following three properties:

I. $\text{Poss}(\emptyset) = 0$

II. $\text{Poss}(W) = 1$

III. $\text{Poss}(p \cup q) = \max(\text{Poss}(p), \text{Poss}(q))$.

Whereas the probability of a proposition is the combined sum of the probabilities of the (exclusive) alternatives with respect to which the proposition is true (that is, according to ordinary probability theory), the possibility value of proposition—for Hamblin, the measure of the proposition’s “plausibility”—is simply **equal** to the possibility value of the **greatest** alternative.

• Observe the axioms entail that
  - at least one element of $W$ receives a possibility value of 1, and (hence)
  - that for every proposition, either it or its negation receives a possibility value of 1.

Note also that it is compatible with the axioms that multiple elements of $W$ receive a possibility value of 1, so that there may be propositions with possibility value 1 whose negations are also 1.

• Call a pair of a set of worlds and a possibility measure over it a **possibility space** $\mathcal{P}$. Now Hamblin’s semantics for probability operators assumes an **epistemic** possibility space, one whose $W$ is restricted to a set of epistemically accessible worlds. For starters we may imagine that, in the unembedded case at least, the relevant epistemic possibility space is determined as a function of context as in Kratzer’s semantics.

• **Truth-conditions.** The preorder expressed by *is at least as likely/probable as* is straightforwardly induced by the contextually determined $\text{Poss}$:

$$p \succeq q \iff \text{Poss}(p) \geq \text{Poss}(q)$$

Like Kratzer, Hamblin defines the one-place operator **probably** to mean ‘more likely than not’: $p$ is probable just in case its possibility value is greater than that of its negation:

$$p \text{ is probable } \iff \text{Poss}(p) > \text{Poss}(\neg p)$$

Various compositional implementations of this abstract model are possible.\(^8\)

---

\(^7\)It is notable, and not widely recognized, that Hamblin [1959] was the first to introduce the concept of a possibility measure, anticipating Zadeh [1978], the contemporary **locus classicus**, by nearly two decades. (Zadeh, like Kratzer, was evidently unaware of Hamblin’s paper.) Hamblin actually used the term ‘plausibility’, not ‘possibility’, for his measures; but I will defer to contemporary usage.

\(^8\)An approach in the broad 2D tradition of Kaplan [1989] for **probably** might be:

$$[\text{probably } \phi]^{c,w,\mathcal{P}(w)} = 1 \text{ iff } \Pr_{\mathcal{P}(w)}(\{w' : [\phi]^{c,w',\mathcal{P}(w')} = 0\}) \neq 1$$

so that the proposition expressed by in a given context $c$ is just:

$$\lambda w. [\text{probably } \phi]^{c,w,\mathcal{P}(w)} = 1$$
2.2 Predictions

- As with Kratzer’s account, a natural way to think of consequence on Hamblin’s semantics is in terms of preservation of truth at a world, given a fixed context. (Alternatively, we could think of consequence as diagonal consequence, or in terms of preservation of “acceptance” with respect to a state of information; it makes no difference.) On that understanding, it is trivial to prove that Hamblin’s semantics validates:

\[(V3.) \text{Probably to not probably not}\]
\[(V4.) \text{Probably-distribution over conjunction}\]
\[(V5.) \text{Disjunction introduction}\]
\[(V6.) \text{Minimality}\]
\[(V7.) \text{Maximality}\]
\[(E1.) \text{Conjunctivitis}\]

Notably, the account also validates:

\[(V11.) \text{Positive Form Transfer}\]
\[(V12.) \text{Complement Transfer}\]

a victory over Kratzer’s account.

- What about these?

\[(V1.) \text{Must to probably}\]
\[(V2.) \text{Probably to might}\]
\[(V8.) \text{Chancy Modus Ponens.}\]
\[(V9.) \text{Chancy Modus Tollens.}\]
\[(V10.) \text{Conditional to comparative.}\]

Whether these are validated depends on the semantics for epistemic modals and indicative conditionals assumed within this model. Hamblin didn’t take a stand here. It is easy to imagine a semantics for epistemic modals which would validate (V1) and (V2). Conditionals, however, raise complications that would take as too far afield; for now I just assume just (V8.)-(V10.) can be covered somehow.

2.3 Problems

- **Collapse of likelihood and certainty.** If a state of information is represented by a possibility space, there appears to be no gap between a proposition’s being certain according to the space and its being merely probable, for the only way that a proposition can be probable on this account is for it to receive a possibility value of 1. So the account cannot handle the most elementary gradations of probability:

(13) Bob is likely to come to the party. But Steve is even more likely.

an absurd result.

**Reply:** replace

\[
p \text{ is probable } \iff \Poss(p) > \Poss(\neg p)\]

with

\[
p \text{ is probable } \iff \Poss(p) > x\]

where \(x\) is a contextually determined threshold.

**Reply to reply:** This move would invalidate *Probably to not probably not.*
• The union property problem again. Like Kratzer’s account, this account validates the union property pattern, and so inherits all of its undesirable consequences.

3 Third semantics: probability spaces

3.1 The theory

• Probability spaces can be defined in various ways. For our purposes, take a probability space \( \mathfrak{P} \) to be a pair \( \langle W, Pr \rangle \) of a set of worlds \( W \) and a function \( Pr \) assigning to each subset of \( W \) a number in \([0,1]\) satisfying the following properties:

I. \( Pr(W) = 1 \)
II. \( Pr(p \cup q) = Pr(p) + Pr(q) \), if \( p \) and \( q \) are disjoint.

Assume the probability space reflects the evidential probabilities. In unembedded contexts, suppose it is determined as a function of context, as in Kratzer’s semantics.

• Naturally we would identify the preorder expressed by \textit{is as at least as likely as} with corresponding preorder by probability:

\[
p \succeq q \text{ iff } Pr(p) \geq Pr(q)\]

Keeping the idea for \textit{probably} from Kratzer and Hamblin, we would say:

\[
p \text{ is probable iff } Pr(p) > .5\]

Various compositional implementations of this abstract model are possible.\(^9\)

3.2 Predictions

• As with Kratzer’s account, we can think of consequence in terms of preservation of truth at a world, given a fixed context. (Alternatively, we could think of consequence in terms of preservation of “acceptance” with respect to a state of information; it makes no difference.) On that understanding, it is trivial to prove that this semantics validates:

(V1.) \textbf{Must to probably}
(V2.) \textbf{Probably to might}\(^10\)
(V3.) \textbf{Probably to not probably not}
(V4.) \textbf{Probably-distribution over conjunction}
(V5.) \textbf{Disjunction introduction}
(V6.) \textbf{Minimality}
(V7.) \textbf{Maximality}
(V11.) \textbf{Positive Form Transfer}
(V12.) \textbf{Complement Transfer}

\(^9\)An approach in the broad 2D tradition of Kaplan [1989] for \textit{probably} would be this:

\[
\lambda w. [\text{PROBABLY } \phi]^{c,w} \Psi(w) = 1 \text{ iff } Pr_{\Psi(w)} \{(w': [\text{PROBABLY } \phi]^{c,w',\Psi(w')} = 1) \} > .5
\]

with the proposition expressed by \textit{probably} \( \phi \) in a given context \( c \) being:

\[
\lambda w. [\text{PROBABLY } \phi]^{c,w} \Psi(w) = 1
\]

\(^{10}\)For (V1.) and (V2) I assume a semantics for epistemic modals in the style of Yalcin [2007b].
Conjunctivitis is invalidated, as is the Union property pattern. And the account does not collapse certainty and likelihood. So this account avoids all of the problems plaguing the accounts of Kratzer and of Hamblin.

- As with Hamblin’s account, we need a separate theory of conditionals in order to assess:

(V8.) Chancy Modus Ponens.
(V9.) Chancy Modus Tollens.
(V10.) Conditional to comparative.

A probability-theoretic analysis of indicative conditionals is beyond the scope of the talk. Plausibly, however, a semantics for indicative conditionals which uses the resources of probability theory can be independently motivated, and a compositional semantics very Kratzer’s restrictor analysis of if-clauses appears to be feasible, or so I have argued elsewhere (Yalcin [2007a], Yalcin [2007b]).

3.3 Digression: integration with the semantics of gradable adjectives

Turning more directly to the detailed compositional semantics:

- Since probable and likely are gradable adjectives and we want as uniform a semantics for the associated degree morphology and modifiers as we can get, it is natural to look to the gradable adjectives literature for direction in sorting out the fine compositional semantics of probability talk (Yalcin [2006], Portner [2009]). Let us briefly consider what happens when we plug in Kennedy [2007].

- Kennedy’s is a scale-based analysis. A scale is a triple \( \langle S, \succeq, D \rangle \) of a set of degrees \( S \), a total order \( \succeq \), and a dimension \( D \). Some examples to give the rough idea:
  - Scale for tall: \( S \) = a set of height magnitudes; a total order \( \geq \) on \( S \) intuitively readable as “is greater than”; the dimension height.
  - Scale for short: \( S \) = a set of height magnitudes; a total order \( \geq \) on \( S \) intuitively readable as “is shorter than”; the dimension height.
  - Scale for probable: from the perspective of a probability-theoretic analysis, \( S \) = the set of possible numerical probabilities (the interval \([0,1]\)), \( \succeq = \geq \), and \( D \) is probability.

Now, Kennedy’s central semantic thesis: gradable adjectives express measure functions. Measure functions are functions from individuals to the degree on the relevant scale to which the individual possesses the property corresponding to the adjective. These measure functions are converted to properties of individual by degree morphology.

This story is built around the comparative case, and makes the superficially morphosyntactically simplest occurrence of gradable adjectives (‘Bob is tall’) the most challenging to analyze. Building on Bartsch and Vennemann [1972], Kennedy proposes a null morpheme pos which supplies the relevant contextually standard for such attributions. How it works for tall:

\[
\begin{align*}
[tall]^c &= \lambda x_v. \text{height}(x) \\
[pos \ tall]^c &= \lambda x_v. \text{height}(x) \succ d_c
\end{align*}
\]

---

11I am thinking, of course, of Adams’s thesis, the thesis that the acceptability of an indicative conditional goes by the corresponding conditional probability (Adams [1965]). It is mostly a commonplace among philosophers that Adams’s thesis is true. See Oberauer and Wilhelm [2003] for psychological evidence broadly supporting Adams’s thesis qua descriptive thesis about how people understand conditional judgments.

12It is a little unclear on Kennedy’s treatment why it is not part of the identity of the set of degrees on a scale that they belong to the dimension that they belong to, so that the dimension parameter is dispensable.

13In English the list includes, inter alia, the comparative morphemes (more, less, as), intensifiers (very, quite, rather, etc.), the usual suffixes (-er, -est) and sufficiency morphemes (too, enough, so).
• Adapting this proposal in the most flatfooted way possible for the language of probability, we have (cf. Yalcin [2006], Portner [2009]):

\[
\begin{align*}
\text{[[probable/likely]]}^c &= \lambda p. Pr(p) \\
\text{[[pos probable/probably/it is likely that]]}^c &= \lambda p. Pr(p) > d_c
\end{align*}
\]

• Remarks and observations:

  – Some type-shifting for the gradable morphology will be required to accommodate the fact that the measure functions corresponding to probable and likely take propositions, not individuals (unless some general nominalization can be claimed).

  – It appears that the relevant probability measures or scales are bindable or shiftable by attitude verbs (Yalcin [2007b]):

    (14) # John imagines that it isn’t raining and that it’s probably raining.

       The embedded probably-claim is not a description of the state of evidence in the context. Indeed it is not a description of any state of evidence.

  – Scales can have or lack maximal elements (be open or closed). Modifiers such as completely appear to be sensitive to such facts; there’s not way to be completely tall, but it is possible for the door to be completely closes. The scale for probable suggested by the probability-theoretic analysis is closed (includes 1). As Portner [2009] observes, however:

    (15) # It is completely probable that Bob is at the party.

       This seems a problem for the probability-theoretic treatment.

       Still, as Kennedy observes, “not all modiers co-occur with all adjectives for apparently idiosyncratic reasons” (34). E.g., the dimension associated with cost has a minimal element (namely, $0$); but it is bad to say, of a free t-shirt, that it is completely inexpensive. So the worry seems not decisive. (It is also possible that there is a competition effect with “it is certain that” which helps explain the infelicity.)

  – Unlike other gradable adjectives, probably and likely don’t enjoy for-PPs:

    (16) Kyle’s car is expensive for a Honda.

    (17) ? Bob’s coming is probable for an event of that sort.

       For-PPs are used in the gradable adjectives literature to argue that the value of $d_c$ can be compositionally manipulated. Are we better off saying that in the case of probable, the relevant standard of comparison is infallibly .5 (as assumed in the probability-theoretic semantics above)? Or should we say it is manipulable, but only by extrasemantic factors? Is there any evidence that the standard for comparison is manipulable by extrasemantic factors?
4 Contrast-sensitivity of probability talk

Some (very preliminary) studies suggest that speaker judgements about the truth of probably $\phi$ are in fact remarkably sensitive to the way that the alternatives to $\phi$ are presented in context.

**Trial 1**

100 subjects

Target question:

A soccer league has several teams competing against each other. Each team has a certain probability of winning, and of not winning, the championship. The chances of a certain team, Team X, are given below.

<table>
<thead>
<tr>
<th>Team X</th>
<th>42 percent chance of winning the championship</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>58 percent chance of not winning the championship</td>
</tr>
</tbody>
</table>

Assuming the above information, is the following sentence true or false?

"Team X will probably win the championship."

Results as percentage of subjects:

- True: 15%
- False: 76%
- Not sure: 9%

**Trial 2**

100 subjects

Target question:

A soccer league has six teams competing against each other. Each team has a certain probability of winning the championship. The teams and their chances of winning are listed below.

<table>
<thead>
<tr>
<th>Team A</th>
<th>12 percent chance of winning the championship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Team B</td>
<td>11 percent chance of winning the championship</td>
</tr>
<tr>
<td>Team C</td>
<td>13 percent chance of winning the championship</td>
</tr>
<tr>
<td>Team D</td>
<td>42 percent chance of winning the championship</td>
</tr>
<tr>
<td>Team E</td>
<td>12 percent chance of winning the championship</td>
</tr>
<tr>
<td>Team F</td>
<td>10 percent chance of winning the championship</td>
</tr>
</tbody>
</table>

Assuming the above information, is the following sentence true or false?

"Team D will probably win the championship."

Results as percentage of subjects:

- True: 76%
- False: 22%
- Not sure: 2%
Comments:

- The surveys were taken separately on virtually disjoint populations of mostly native speakers of English.

- Note, subjects are not asked to estimate a probability (as in, say, the familiar conjunction and base rate fallacies); they are simply given the probability of the outcome in question. The point of these surveys is to understand how (if at all) the truth of probably-φ-judgements depend on the exact probabilities.¹⁴

- The first survey is the control. Team X of survey 1 “reappears” as Team D in survey 2. Even though the event of Team D/X’s winning is the same in each survey (42%), the judgments about the truth of “Team D will probably win” completely reverse between survey 1 and survey 2. Prima facie this suggests that the truth of “Team D will probably win” is sensitive to the way the alternatives to the prejacent are partitioned in context.

- It is difficult to make semantic sense of this data.

1. Could we try saying that probably φ means something like “φ is (much) more likely than the relevant alternatives”? Consider a lottery of a million tickets. I buy 300 tickets; everyone else buys just one. So I am much more likely (300 times more likely) than anyone else to

¹⁴Note: it would have been preferable to set up the trials without using technical language like “percent chance”. A new study using simpler language is under way.
win the lottery. But it is very hard to believe people would say, in this case, that I will probably win the lottery. So this approach does not seem plausible.

2. We could try saying that probably φ means something like “φ is (much) more likely than contextually set threshold $x^*$, and that $x$ may take a value below .5. But this would mean that in some contexts, both φ and $\neg\phi$ are probable, meaning that the inference pattern **Probably to not probably not** is not really valid. This is hard to believe. (And preliminary studies suggest that subjects will in fact judge that “Team D will probably not win the championship” is false in the Trial 2-style presentation of the data, a judgment that preserves (V3).)

- If the data stand up, this appears to be a case of **superadditivity** (Cf. Macchi et al. [1999]).

---

**References**


Seth Yalcin. A paradox of epistemic modality. Notes for a talk to Department of Linguistics, University of Massachusetts, Amherst., 2006.

