Philosophy Compass 5/11 (2010): 916–937, 10.1111/j.1747-9991.2010.00360.x

Probability Operators

Seth Yalcin*
University of California

Abstract
This is a study in the meaning of natural language *probability operators*, sentential operators such as *probably* and *likely*. We ask what sort of formal structure is required to model the logic and semantics of these operators. Along the way we investigate their deep connections to indicative conditionals and epistemic modals, probe their scalar structure, observe their sensitivity to contextually salient contrasts, and explore some of their scopal idiosyncrasies.

1. *Introduction*

This is about the meanings of *probability operators*. Take as (English) paradigms those sentential operators constructable from *probable* and *likely*—for instance, *probably*, *it is likely that*, and the various matching two-place comparative operators (*is more likely than*, *is as probable as*, etc.). Our question is how best to model the meanings these expressions have in natural language.

The question is of some interest from a number of directions. At the level of compositional semantics, probability operators present some novel and intriguing obstacles to straightforward analysis, and getting their treatment right is likely to shed light on important allied constructions, such as epistemic modals, indicative conditionals, and gradable adjectives. Second and related, probability talk raises questions at the interface of compositional semantics with pragmatics. I obviously communicate some information when I say 'It is probably raining'. What kind of information is this, and how should it be modeled? Is some truth-conditional content expressed and communicated with this sentence, or should we understand its communicative impact differently? There are reasons to suspect that probability talk recommends distinctive answers to these questions. Third, a clearer view about how probability talk is used and interpreted is relevant to the assessment of experimental work in psychology relying on the use of such talk by subjects. Probability talk exhibits context-sensitivity, and the detailed nature of this context-sensitivity may have an important effect on the space of possible explanations for why subjects might make or fail to make a given probability judgment. Fourth, working through the semantics and pragmatics of probability operators may be clarifying for philosophical analyses of various concepts of probability, as it arguably has been for philosophical inquiry into modality.

This paper is devoted to the matter of the semantics of probability operators. We will barely graze questions about the pragmatic impact of probability sentences on context, issues in metasemantics, connections with work in psychology (though see section 6), and connections with philosophical inquiry into the nature of probability. Take this paper as a semantic preliminary to such investigations.

Note that our question is distinct from the (related) question of how the tools of probability theory might be profitably applied to diverse fragments of language. On that, see Cohen (2009).

2. Background Assumptions

It is common to theorize about the semantics of probability operators under the following assumptions. First, *probable* and *likely* are, or serve chiefly to produce, sentential operators. We approach their compositional semantics from this perspective. They semantically combine with sentence meanings—either directly, or after first combining with some further, perhaps hidden, morphology. Second, *probable* and *likely* make the same semantic contribution; what distributional differences they have are explained by their non-semantic features. Third, all probability talk is fundamentally *comparative* in nature. This assumption recommends a strategy for theorizing about these operators: begin with an analysis of the comparative forms (*is as probable as*, *is more likely than*, etc.), and then analyze the superficially one-place operators (*probably*, *it is unlikely that*, etc.) as covert comparatives. In following this kind of strategy, work on the semantics of probability talk parallels the gradable adjectives literature, where it is fairly standard to take the comparative form as expressing the 'theoretically basic' notion.[1] The parallel here is not surprising, since *probable* and *likely* really just are gradable adjectives.[2] I will follow the literature, and make all of these assumptions.

Our project is continuous with compositional semantic inquiry in general. But there is another, rather different kind of project about ''interpretation'' in our vicinity, one that sometimes gets called ''the interpretation of probability'' (for an overview, see Mellor (2005), Hajek (2009)). This sort of project is largely concerned with the following metaphysical question: which aspects of reality have (something approaching) the structure of a mathematical probability space, and in virtue of what do those aspects of reality have that structure? Some remarks are in order about the relation of the present paper to this body of work.

A number of interpretations of probability have been investigated; most fit at least roughly into one of the following three camps. First there are *objective chance* interpretations, according to which features of the world itself are taken to have a probabilistic structure. (Perhaps owing to the dispositions and propensities of the things in the world, as on a propensity view; perhaps owing to the frequencies with which events occur, as on a frequentist view.) Second, there are *evidential* interpretations of probability, wherein the relation of confirmation or evidential support between propositions is taken to have a probabilistically articulated structure. On one way of developing this idea, the probabilities reflect 'something like the intrinsic plausibility of hypotheses prior to investigation' (Williamson 2000: 211). Third, there is the *subjective Bayesian* view that the attitude state of belief (confidence, expectation, doubt) has a structure approximating that of a probability space. Here probabilities correspond to degrees of belief, or credences.

One might accept that all of these domains have a probabilistic structure, with none reducible to the other; or one might argue that one of these interpretations is more fundamental, and serves to ground the others.

Philosophers familiar with the debates surrounding these different views of probability may be tempted to ask: which sort of probability is in focus in this paper? The question misconstrues the enterprise. We will consider natural language as we find it, without making assumptions about the nature of the domain(s) being described in advance. In fact, we will not even assume in advance that probability operators are to be interpreted via recourse to a probability space. That is a semantic thesis, and it requires semantic evidence. As we will see, one can make considerable progress limning the logic and compositional structure of probability operators in abstraction from substantive metaphysical assumptions.

Still, suppose we could adequately motivate, on semantic grounds, the idea that probability operators are to be interpreted via recourse to a probability space. (I will supply such motivation below.) Another question one might then ask is whether a particular interpretation, or class of interpretations, of probability is best suited to the content of natural language probability claims.

That is an interesting open question. Philosophers and linguists are familiar with treatments of modality which allow an unembedded modal sentence to express various flavors of relative necessity and possibility, depending on context (see Kratzer (1991), and section 3.1 below). From the perspective of such accounts, it is natural to suspect that a sentence with a wide-scoped probability operator could allow for a similar kind of relativity, allowing various 'flavors' (interpretations) of probability depending on context. The various semantic proposals I discuss below may in principle be made compatible with this kind of variability. Our main focus, however, is on matters at a higher level of abstraction. Any precise version of a view along these lines will ultimately need to be expressed in terms of the compositional semantics of probability operators; and that semantics will have to recon with their fundamental logical interactions, their scoping and embedding properties, and the like. It is these matters that will take up most of our attention.

3. The Relative Likelihood Approach: Kratzer

3.1. OVERVIEW

The first proposal we consider takes it that ordinary probability talk involves only a qualitative, not quantitative, notion of likelihood—that it adverts only to a relative ordering of propositions, leaving numerical magnitudes of probability out of the story. I will call this the *relative likelihood* approach. A proponent of this approach might take her relative likelihood ordering on propositions as primitive, or she might define it in terms of other primitives. We consider one way of doing the latter, due to Kratzer (1991).[3]

Kratzer groups probability operators semantically with modals, and so her account is situated within her treatment of modality generally. On this approach, modal expressions (*might, can, ought, should, must, possibly*, etc.) are analyzed in the metalanguage as effecting quantification over possible worlds relative to two contextually supplied parameters. One parameter—the *modal base*—delivers the domain of modal quantification. 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. So, probability claims will say something about how things stand with respect to some body of evidence. The second parameter—the *ordering source*—imposes a reflexive, transitive relation (preorder) $\geq$ over the worlds supplied by the modal base. As for the interpretation of probability operators, Kratzer takes the worlds to be preordered by 'normality' or 'stereotypicality' (Kratzer 1991: 644). The fact that some modals (*may, can, ought, should, have to, must*, etc.) allow for a multiplicity of interpretations (deontic, epistemic, nomological, etc.) is explained by appeal to permissible variability in the modal base and ordering source parameters; see von Fintel and Gillies (2007), Swanson (2008) for more on Kratzer's treatment of modality in general.

How does Kratzer get her preordering of worlds in terms of normality? The ordering source delivers, at the world of evaluation, a set of O of propositions, a set of propositions *normally true* in the relevant situation. (Propositions are construed as sets of worlds.) This set is then used to induce a preorder on worlds as follows:

$$w \geq_O w' \quad \text{iff } \{p \in O : w \in p\} \supseteq \{p \in O : w' \in p\}$$

If two worlds w, w' make just the same 'normal' propositions from O true, then $w =_O w'$. If w makes all the same propositions from O true as does w' *and then some*, then $w >_O w'$. Finally, if w and w' each makes a proposition from O true that the other does not, then the two worlds will be incommensurable by $>_O$.

Now the English comparative construction *is at least as as probable/likely as* semantically expresses some reflexive, transitive relation $\succsim$. Kratzer uses the normality preorder ≥ over worlds supplied by the ordering source to define this relation and (thereby) to give the compositional semantic contribution of *is at least as as probable/likely as* ($\succsim$). Here is the definition:

$$p \succsim q \quad \text{iff } \forall w \in q : \exists w' \in p : w' \geq w$$

(Understand our quantifiers as restricted already by the relevant epistemic modal base.) From this treatment, the semantics for other comparatives falls out. A proposition *p is as likely/probable as q* ($p \sim q$) 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, and *vice versa*. The strict comparative *is more likely than* ($\succ$) is then defined in terms of $\succsim$ as follows:

$$p \succ q \quad \text{iff } (p \succsim q) \wedge \neg(q \succsim p)$$

Finally, the one-place operator *probably* is defined in terms of $\succ$ to mean 'is more likely than not'. (For instance, 'It is probably raining' is true just in case it is more likely to be raining than not raining.) Kratzer does not attempt to unpack the internal semantics of the comparative forms.

One might ask whether the relevant modal bases, ordering sources, accessibility relations, etc., are syntactically represented, or are referred to only in the semantic metalanguage. The formalism of Kratzer (1991) assumes the latter view, though the general picture might be developed in either direction.

Recapping: first, context supplies a set of epistemically accessible worlds and a set of propositions to serve as the ordering source; second, the ordering source induces a preorder over the epistemically accessible worlds; third, this preorder on worlds induces the preorder $\succsim$ on propositions semantically expressed by *is as at least as probable/likely as*; finally, semantics for the rest of the probability operators are defined in terms of $\succsim$. On this account, probability operators are effectively gradable epistemic modal operators.

3.2. ADVANTAGES OF THE RELATIVE LIKELIHOOD ACCOUNT

A difficulty in developing an intuitive picture of Kratzer's account is that it is often not obvious in particular cases how comparative judgments of probability are to be traced back to intuitive constraints on the relevant ordering source. Part of the issue may be that which propositions are normal in a situation depends heavily on just which features of the situation are held fixed, and no recipe has been offered for settling, in a given context, what to hold fixed. Another issue may be that the ordinary notions of normality and of probability differ in important ways. One can say that Bob is *probably* not in his office, though he *normally* is; to say that bears *normally* hibernate in winter is quite different from saying that bears *probably* hibernate in winter; and so on. Nevertheless, we can get a better sense of what Kratzer's account amounts to by examining the range of inference patterns it validates.[4]

Kratzer's semantics validates a wide range of intuitively valid patterns. Consider first some basic interactions of *probably* with logical operators (assuming the standard boolean semantics for these operators):[5]

V1. *Probably* to *not probably not*.

$$\frac{\text{probably } \phi}{\neg \text{ probably } \neg\phi}$$

V2. Distribution over conjunction.

$$\frac{\text{probably } (\phi \wedge \psi)}{\text{probably } \phi \wedge \text{ probably } \psi}$$

V3. Chancy disjunction introduction.

$$\frac{\text{probably } \phi}{\text{probably } (\phi \vee \psi)}$$

I take it these are all intuitively valid. Kratzer's account has them covered. (Proofs are elementary and will generally be left to the reader.)

Next consider:

V4. Minimality

$$\frac{}{\phi \text{ is at least as likely as } \bot}$$

V5. Maximality

$$\frac{}{\top \text{ is at least as likely as } \phi}$$

These are rather more abstract and farther removed from ordinary linguistic judgment, but I take it we want this patterns validated: surely a tautology is at least as likely as any proposition you please, and surely any proposition you please is at least as likely as a contradiction. Kratzer's account has these covered as well.

We get further advantages from Kratzer's semantics when we combine it with her semantics for epistemic modals (see Kratzer (1991) for the details). In particular, we cover:

V6. *Must* to *probably*.

$$\frac{\text{It must be that } \phi}{\text{Probably } \phi}$$

V7. *Probably* to *might*.

$$\frac{\text{Probably } \phi}{\text{It might be that } \phi}$$

which are both intuitively desirable. (The idea that *probably* is something like 'intermediate in strength' between epistemic *must* and *might* was part of Kratzer's motivation for classifying probability operators with the epistemic modals in the first place.)

We get still more advantages when we also accept Kratzer's semantics for indicative conditionals, which integrates elegantly with her account of modality. Briefly, the idea is that indicative conditionals are modalized sentences, and that *if*-clauses are devices for introducing an additional restriction on the modal base of the relevant modal operator, a restriction to worlds in which the antecedent is true.[6] (Where no modal or probability operator is superficially apparent, a tacit epistemic necessity modal is assumed. For more details, see Kratzer (1986), and Section 3.3 below.) With this semantics assumed, Kratzer secures:

V8. Chancy Modus Ponens.

$$\frac{\begin{array}{l}\text{if } \phi \text{ then } \psi \\ \text{probably } \phi\end{array}}{\text{probably } \psi}$$

V9. Chancy Modus Tollens.

$$\frac{\begin{array}{l}\text{if } \phi \text{ then } \psi \\ \neg \text{ probably } \psi\end{array}}{\neg \text{ probably } \phi}$$

V10. Conditional to comparative.

$$\frac{\text{if } \phi, \text{ then } \psi}{\psi \text{ is at least as likely as } \phi}$$

I take it (V8) and (V9) are intuitively obvious, but Conditional to comparative might seem surprising. However, 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, 'ϕ is more likely than ψ' entails 'it might be that $(\phi \wedge \neg\psi)$', which (in a fixed context) seems intuitively incompatible with 'if ϕ, then ψ'. So, this seems like a pattern we do want to cover, and Kratzer's package validates it.

3.3. SHORTCOMINGS OF THE RELATIVE LIKELIHOOD ACCOUNT

These are impressive results. Helping itself only to quantification over some preordered worlds, Kratzer's semantics nevertheless makes surprisingly fine distinctions and covers a wide range of inference patterns. It also coheres naturally with an independently motivated and powerful semantics for modals and conditionals. Unfortunately, however, the relative likelihood model of probability underlying this account has severe shortcomings. There are valid inference patterns that Kratzer's account fails to validate, and there are patterns the account validates which are clearly invalid.

Beginning with inferences of the first sort, consider a pattern we might call:

V11. Positive Form Transfer[7]

$$\frac{\begin{array}{l}\psi \text{ is at least as likely as } \phi \\ \text{Probably } \phi\end{array}}{\text{Probably } \psi}$$

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 equally valid:

V12. Complement Transfer

$$\frac{\begin{array}{c}\psi \text{ is at least as likely as } \phi \\ \phi \text{ is at least as likely as } \neg\phi\end{array}}{\psi \text{ is at least as likely as } \neg\psi}$$

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 extremely plausible. (If you are not convinced, try breaking the pattern into two easier-to-process patterns.) Neither of these inference patterns are valid on Kratzer's semantics. I sketch a countermodel in a note.[8]

Moving to invalid inference patterns validated by the account, it is known that the relative likelihood model assumed by Kratzer validates:

I1. Union property pattern.

$$\frac{\begin{array}{c}\phi \text{ is at least as likely as } \psi \\ \phi \text{ is at least as likely as } \chi\end{array}}{\phi \text{ is at least as likely as } (\psi \vee \chi)}$$

(See Halpern 1997, 2003.) This is an egregious result; after all,

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.

appears to be obviously invalid reasoning.

One might see room to maneuver: perhaps the conclusion here somehow mandates a wide scope *or* interpretation, making this reasoning a non-instance of (I1). But the underlying problem can be stated out without having to nail the semantics of *or*. A special case of (I1) is:

$$\frac{\begin{array}{c}\phi \text{ is as likely as } \phi \\ \phi \text{ is as likely as } \neg\phi\end{array}}{\phi \text{ is at least as likely as } \top}$$

Notice two things here. First, the first premise of this pattern is trivially true and (hence) can be deleted. Second, observe that if ϕ is at least as likely as $\top$, then ϕ is at least as likely as any proposition you please by (V5) (Maximality). This entails the following inference pattern is validated:

I2. Collapse of equiprobability into certainty.

$$\frac{\phi \text{ is as likely as } \neg\phi}{\phi \text{ is at least as likely as } \psi}$$

So on the relative likelihood semantics, if the coin's landing heads is as likely as its landing tails (not heads), it is as likely as any proposition you please.

Reductio. I take it these failures tell decisively against Kratzer's relative likelihood semantics.

We might wonder, however, whether something within the general vicinity can be salvaged. Is there some better way of extending a preorder over worlds to a preorder over propositions, one which will get the inference patterns right? See Halpern (2003) for some interesting other ways to produce a preorder on propositions from a preorder on worlds. While I have no general impossibility proof, there is the obvious worry that such non-numerical, qualitative approaches to probability talk will inevitably leave out overtly quantitive uses of this talk. For instance:

(1) It's .5 likely that the box contains a thousand dollars.
(2) There is a 60% probability that it's raining. (Portner 2009: 73)

It's unclear how such talk is to be accounted for in a purely qualitative semantics.

One might be tempted to classify these cases as employing a technical use, one discontinuous with ordinary natural language probability talk. I have some sympathy with this response. But while it is of course true that we should not be misled by stipulatively defined technical vocabulary, we should also remember that claims like (2) entail various quite ordinary probability claims (for instance, 'It's not likely it will not rain', and 'It might be raining'). Giving them a wholly autonomous semantics may make predicting these valid patterns of inference more difficult. Moreover, there are less technical-sounding, but still superficially quantitative, uses of this talk:

(3) The coin is twice as likely to land heads as tails.

Such sentences are used and interpreted by speakers who are innocent of the mathematical theory of probability. It is clear that a semantics for probability operators must reckon with them.

Finally, as noted already above (note 4), probability operators are gradable adjectives, and leading contemporary accounts of the semantics of gradable predicates models them as relations to degrees (see, e.g., Kennedy 2007). Degrees need not correspond to numerical magnitudes, but in the case of predicates that correspond to intuitively measurable properties, they often do. It seems natural, then, to explore the possibility of a semantic account of probability operators that helps itself to quantities. Hamblin (1959), which we consider next, is one such account.

4. The Possibility Space Approach: Hamblin

It is interesting that Kratzer does not seriously consider the question of using a conventional probability model in the semantics of probability operators. By contrast, Hamblin (1959) opens with that question:

> Metrical probability theory is well-established, scientifically important and, in essentials, 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? (Hamblin 1959: 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, following Keynes, 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).

Aiming to model the spirit of this kind of view, Hamblin defined a new kind of measure for representing uncertainty, one less restrictive in important respects than that of ordinary probability. The class of measures he defined are nowadays called *possibility measures*.[9] Suppose for simplicity that W is a finite set of worlds. A possibility measure *Poss* is any function assigning to each subset of W a number in [0, 1] satisfying the following three properties:

I. $Poss(\emptyset) = 0$
II. $Poss(W) = 1$
III. $Poss(U \cup V) = \max(Poss(U), Poss(V))$.

Whereas the probability of a proposition is the combined sum of the probabilities of the alternatives with respect to which the proposition is true, 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 that the axioms entail that at least one element of W receives a possibility of 1, and (hence) that for every proposition, either it or its negation receives a possibility 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*. Now Hamblin's semantics for probability operators assumes an *epistemic* possibility space, one whose W is restricted to a set of epistemically accessible worlds. We can implement Hamblin's approach semantically in a manner parallel to Kratzer's account. Recall on her account, context supplies some functions which eventually give us, at an evaluation world, a set of epistemically accessible worlds preordered by something like normality. In the current setting we can say something similar: a context c supplies a function f^c from worlds to an epistemic possibility space.[10]

Now the preorder over propositions expressed by *is at least as likely/probable as* is straightforwardly induced by the *Poss* function given by the possibility space determined at the evaluation world:

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

The comparatives *is as likely as* and *is more likely than* fall out in the obvious way. 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:

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

As with Kratzer's account, the natural way to think of consequence on Hamblin's semantics is in terms of preservation of truth at a world, given a fixed context. On that understanding, it is trivial to prove that Hamblin's semantics validates (V1)–(V5): Probably to not probably not, Probably-distribution over conjunction, Disjunction introduction, Minimality, and Maximality.

Whether inference patterns (V6)–(V10) are validated in Hamblin's semantics depends on the semantics of epistemic modals and of conditionals assumed. We must understand epistemic modal and indicative conditional sentences as imposing conditions on possibility spaces for the patterns to come out valid. In the case of epistemic possibility and necessity modals, a natural thing to say is that these are quantifiers (existential and universal, respectively) over the worlds given by the possibility space supplied by the evaluation world. This would validate *Must* to *probably* (V6) and *Probably* to *might* (V7). Extending Hamblin's semantics to conditionals is a more delicate matter; I will not attempt that here, and so will set consideration of (V8)–(V10) aside.

Notably, Hamblin's semantics validates (V11) (Positive Form Transfer) and (V12) (Complement Transfer), a significant point over Kratzer's account. However, it also validates the following eyebrow-raising pattern, which I will call, nodding to Kyburg (1970), Conjunctivitis:

E1. Conjunctivitis.

$$\frac{\begin{array}{c}\text{probably } \phi \\ \text{probably } \psi\end{array}}{\text{probably } (\phi \wedge \psi)}$$

Kratzer's account, by contrast, does not validate this pattern.[11,12]

Is the pattern valid or invalid? Very often, this inference sounds fine, Hamblin (1959) observes. Indeed, it is not easy to come up with completely natural-sounding counterexamples. But of course, it is easy to imagine the sort of cases which *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 12-sided die, with the sides numbered 1–12. 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:

(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.[13] The upshot of this is that the tendency to modally subordinate makes it hard to probe the validity of this inference pattern directly. Absent further considerations, we should allow theoretical considerations to settle the question whether this pattern is valid, and not take its validity as an explanandum.

In any case, Hamblin's account stumbles over far clearer cases. Like Kratzer's account, Hamblin's semantics validates (I1) Union property pattern, together with its absurd consequence, repeated here:

I2. Collapse of equiprobability into certainty.

$$\frac{\phi \text{ is as likely as } \neg\phi}{\phi \text{ is at least as likely as } \psi}$$

It is easy to see why. Recall that every proposition or its negation receives a possibility value of 1. This means that if ϕ is just as likely as $\neg\phi$, then *both* ϕ and its negation have the possibility value 1. And since this is the maximal value, we can conclude that ϕ is at least as likely as any proposition you please.

Note also that from the fact that every proposition or its negation receives a possibility value of 1, and from the semantic assumption that *probably* ϕ is true just in case ϕ has a greater possibility value than its negation, it follows Hamblin validates:

I3. Hamblin's collapse.

$$\frac{\text{Probably } \phi}{\phi \text{ is at least as likely as } \psi}$$

—for the only way for a proposition to be probable is for it to receive a possibility value of 1. So judgments like:

It's likely that Bob is at the party, but it's even more likely that Steve is at the party.

make no sense on Hamblin's system.

These are absurd results, and there is no easy fix. The possibility space model Hamblin discovered has powerful applications (as the work building on Zadeh (1978) has demonstrated), but it is doubtful that one of those applications is to the semantics of probability operators.

5. Semantics with Probability Spaces

Let us consider, finally, using standard probability theory in the semantics of probability operators. Versions of this approach has been investigated recently by Swanson (2006) and Yalcin (2007) (see also Portner 2009, Swanson, forthcoming).

Kratzer's account relativized extensions to modal base functions f from worlds to sets of worlds. We can easily upgrade this account probabilistically. Rather than relativizing extensions to functions from worlds to sets of worlds, let us relativize extensions to functions e from worlds to probability spaces. Probability spaces can be defined in various ways, depending on the character of the sample space and depending on which notions are taken as primitive (see Hájek (2003, 2009) for discussion). To avoid irrelevant technicalities, let us suppose that the space of all possible worlds W is finite. An *epistemic probability space* is a pair $\langle E, Pr\rangle$ of a set of worlds E (some subset of W) and a function Pr assigning to each subset of W a number in [0, 1] satisfying the following:

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

Semantically, the story we will tell parallels the story we told in the possibility space setting, which in turn parallels Kratzer's story. When we interpret probability operators, we assume the context c supplies a function g^c from worlds to probability spaces. (Extending Kratzer's terminology, we might call it a 'probabilistic conversational background'.) The probability space, we take it, captures some particular body of evidence determined as a function of the evaluation world. Effectively, it plays the role that Kratzer's preordered epistemic modal bases did. We assume that E corresponds to a set of worlds epistemically accessible from the evaluation world. All of the 'probability mass' is located within E.

Now the preorder over propositions expressed by *is at least as likely/probable as* is straightforwardly induced by the Pr function given by the probability space determined at the evaluation world:

$$p \succsim q \quad \text{iff } Pr(p) \geq Pr(q)$$

The comparatives *is as likely as* and *is more likely than* fall out in the obvious way. If we want, we can follow Hamblin and Kratzer when it comes to *probably*, taking it to mean 'more likely than not':

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

which of course is equivalent to:

$$p \text{ is probable} \quad \text{iff } Pr(p) > .5$$

To anticipate our need for semantic interaction between probability operators and epistemic modals, we can provide a semantics for epistemic modals within this setting. If E is the set of epistemically accessible worlds fixed by the relevant epistemic probability space at the evaluation world,

$$p \text{ is epistemically possible} \quad \text{iff } \exists w \in E : p(w) = 1$$
$$p \text{ is epistemically necessary} \quad \text{iff } \forall w \in E : p(w) = 1$$

This allows us to maintain the idea that the 'pure' epistemic modals are quantificational in character—a desirable result, one in line with standard views about epistemic modals.

In more formal attire, we have:

$$[[\text{MIGHT}\phi]]^{w,e} = 1 \quad \text{iff } \exists w' \in E_{e(w)} : [[\phi]]^{w',e} = 1$$
$$[[\text{MUST}\phi]]^{w,e} = 1 \quad \text{iff } \forall w' \in E_{e(w)} : [[\phi]]^{w',e} = 1$$
$$[[\text{PROBABLY}\phi]]^{w,e} = 1 \quad \text{iff } Pr_{e(w)}(\{w' : [[\phi]]^{w',e} = 1\}) > .5$$

We can now note some results. This semantics validates (V1)–(V7): *Probably* to *not probably not*, Probably-distribution over conjunction, Disjunction introduction, Minimality, Maximality, *Must* to *probably* and *Probably* to *might*. It validates Positive Form Transfer (V11) and Complement Transfer (V12), a point over Kratzer's account. It also invalidates the Union property pattern (I1), escaping its absurd consequences, a significant point over both Kratzer's and Hamblin's accounts. Finally, we note the account is free of Conjunctivis (E1).

So far, so good. It remains to assess the patterns (V8)–(V10). To do this we must make some assumptions about the semantics of indicative conditionals.

I offer the following probabilistic modification of Kratzer's semantics for conditionals (Kratzer (1986, 1991); related developments include Yalcin (2007), Kolodny and MacFarlane (2010)). Again, Kratzer's idea, building on Lewis, is that conditional meanings are the result of the separate semantic contributions of an *if*-clause and a (perhaps tacit) modal operator, with the role of the former to restrict the quantificational domain of the latter. On my proposed upgrade, *if*-clauses will still perform the semantic function they performed on Kratzer's account, but they will do more besides. These clauses will now shift *e* functions rather than modal base functions, as follows:

$$[[(\text{IF}\phi)\psi]]^{w,e} = [[\psi]]^{w,e^{\phi}}$$

where e^{ϕ} comes from e and the meaning of the antecedent clause ϕ as follows: if $e(w) = \langle E, Pr\rangle$, then $e^{\phi}(w) =_{DEF} \langle E^{\phi}, Pr^{\phi}\rangle$, where E^{ϕ} is defined as the set of ϕ-worlds in E,[14] and where Pr^{ϕ} is defined as *Pr conditionalized* on ϕ.[15] (We assume, in the style of Kratzer (1991), that the clause in the position of ψ is always a modal claim.)[16]

The basic idea here is simple. What indicative *if*-clauses did before was restrict the quantificational domain for an epistemic modal. They will still do that here, by shifting the sample space E given by $e(w)$ to the subset of that space wherein the antecedent is true. But now they do one more thing: they shift the probability measure supplied by $e(w)$, by conditionalizing on the antecedent. One rough way to think of it: the *if*-clause shifts e to the nearest epistemic probability space that includes the information in the antecedent. Another: it shifts e to the minimal revision of e that would include the antecedent (albeit adopting a substantive view here about what 'minimal' amounts to).

Observe that conditionals which don't involve probability operators—e.g., 'If the lights are on, Bob must be in his office'—are left unharmed by this analysis. Schematically, the semantics is parallel to Kratzer's:

$$[[(\text{IF}\phi)(\text{MUST}\psi)]]^{w,e} = 1 \quad \text{iff } \forall w' \in E_{e^{\phi}(w)} : [[\psi]]^{w',e^{\phi}} = 1$$

This is because epistemic necessity modals do not check the probability function component of e for truth; they care only for $E_{e(w)}$, which plays just the role of Kratzer's epistemic modal bases. The probability function component is what is consulted when interpreting—surprise—probability operators. A sentence like 'If the lights are on, Bob is probably in his office' is rendered schematically as

$$[[(\text{IF}\phi)(\text{PROBABLY}\psi)]]^{w,e} = 1 \quad \text{iff } Pr_{e^{\phi}(w)}(\{w' : [[\psi]]^{w',e^{\phi}} = 1\}) > .5$$

The conditional is true just in case the probability that Bob is in his office, conditional on the lights being on, is sufficiently high. We have, then, a fairly conservative probabilistic extension of a powerful existing conditional semantics, one (*pace* Lewis (1976)) forging a semantic link between indicative conditionals and conditional probabilities.[17]

This is not the place to defend this analysis of indicative conditionals in any detail. I settle for merely reporting the results of this semantics as concerns the inference patterns Chancy Modus Ponens (V8), Chancy Modus Tollens (V9), and Conditional to comparative (V10). Results: we validate all of these patterns.

The probability space semantics defeats the relative likelihood and possibility space accounts as concerns all of the desiderata we have so far considered. (It is an interesting and worthwhile question, one to my knowledge open, whether a semantics exploiting a different underlying representation of uncertainty—for instance, one using Dempster-Shafter belief functions (see Shafer 1976), the support theory of Tversky and Koehler (1994), or sets of probability measures—can do the job equally well.) It is time to search for more desiderata. We consider three further sets of issues.

6. Scales and Alternatives

As noted, in the literature on comparative constructions, the leading views model gradable predicates as expressing relations to *degrees*, intuitively understood as non-relative, objective magnitudes to which an object might possess a property. On this model gradable predicates are associated with a *scale* of degrees, where a scale is defined to be a triple of a set of degrees, a total order on that set, and the property or *dimension* along which the degrees vary. For instance (and glossing over many complications), *tall* is associated with a scale which orders the space of magnitudes of height (the dimension) by the *is greater than* relation. Then if I say

(4) Adam is taller than Heather

the gloss in our semantic metalanguage will be that what I say is true just in case Adam's degree of height is greater than Heather's degree of height. The adjective *short* involves the same dimension, but with the reverse ordering of the degrees (hence a different scale). This is why (4) entails

(5) Heather is shorter than Adam

Appeal to scales helps us to cover these and related entailments, and to do so in a manner that allows a uniform semantic treatment of degree morphology (which in English includes morphemes like *less, as, too, enough, so, how, more -er*, and much else). For instance, we clearly want to say that the suffix *-er* makes a single compositional contribution to *taller* and *shorter*, respectively. There are straightforward paths to doing this if we suppose that *-er* picks up on the strict order supplied by the scale of the adjective it combines with.

Now we can ask what is suggested about the semantics of *probable* and *likely* qua gradable adjectives, if the broadly scalar approaches to gradable adjectives are on track. As it is impossible to wade seriously into contemporary debates about gradable adjectives here, I will just make one observation.

First, there is a distinction between *absolute* and *relative* gradable adjectives (Kennedy 2007). The latter, but not the former, have variable and apparently context-dependent interpretations in unmarked uses. Examples of the latter are *tall, short, expensive, large*; examples of the former are *wet, pure, open, straight*. While adjectives of both classes appear in comparative constructions, unmarked appearances of adverbs in the absolute class appear to relate their objects to fixed positions on the relevant scale (in particular, the maximum or minimum values). For example:

(6) The door is open. [analysis: aperture(the door) > min(aperture)]

This sentence is (strictly speaking) true, it seems, just in case the door has a non-zero degree of aperture. Contrast with

(7) Adam is tall. [analysis: height(Adam) > h_c]

Whether this is (strictly speaking) true depends on further background assumptions, such as what kind of person Adam is—toddler or basketball player?—and what height-bearing entities are contextually relevant. There is significant debate as to exactly what factors contribute to the determination of the comparison value h_c, but the main point for our purposes is that *tall* exhibits a certain relativity that *open* lacks.

If we accept this distinction among gradable adjectives, the question arises whether *likely* and *probable* are absolute or relative. The tacit assumption of the probability space semantics given above is that they are absolute, expressing relations between propositions and a contextually invariant probability, namely .5.[18] Recall this was motivated by the idea, built already into the work of both Kratzer and Hamblin, that a proposition is probable if and only if it is more likely than not. Now, the left–right direction of this biconditional, which corresponds to inference pattern (V1), seems unassailable (though we will assail it in a moment): how could a proposition be probable, if its negation is more likely?

But the right–left direction is more questionable. Consider a fair coin biased with a thin piece of tape, so that it is just slightly more likely to land heads than tails. It is flipped. It seems wrong to say:

(8) The coin probably landed heads.

for (8) tends to suggest that the coin's landing heads is *appreciably* more likely than not, and it isn't. The probability space semantics does not predict this judgment. Absent some pragmatic explanation, we have motivation for looking into bringing *probably* and *likely* closer to the paradigm of relative adjectives.

Let me mention three ways of upgrading the probability space semantics with additional relativity. First, one could replace the *greater than* relation (>) used in the semantics with the *appreciably greater than* relation ($>_c$), where what counts as *appreciably greater than* is a function of > and context:

$$p \text{ is probable} \quad \text{iff } Pr(p) >_c Pr(\neg p)$$

This would suffice to predict (8)-like judgments. Second, one could propose that the relevant comparison probability value n is variable and context-sensitive, and is computed in whatever way it is computed for relative gradable adjectives:

$$p \text{ is probable} \quad \text{iff } Pr(p) > n_c$$

Third, one could assume both kinds of relativity:

$$p \text{ is probable} \quad \text{iff } Pr(p) >_c n_c$$

The second/third options essentially classify *probable* and *likely* as standard relative adjectives. The central point to note in connection with these options is that, unless more is said, they leave open the possibility that n_c could take a value *below* .5. This would put the pattern (V1), *Probably* to *not probably not*, in question, a surprising result.

In fact, there is clear empirical evidence that subjects will assent to the truth of a *probably* φ claim even when given the information that the probability of φ is *less* than half. These are cases where φ has a contextually salient class of alternatives, and the individual alternatives each have a probability appreciably less than that of φ. To illustrate, consider the sentence:

(9) Bloggs is probably the winner of the lottery.

Suppose the lottery in question is a fair one, with a single ticket chosen as the winner out of 1000 possible tickets. Now consider the following two possible sets of additional background facts:

> Background *A*: Bloggs has 420 tickets, and another player, Smith, has the remaining 580 tickets.
> Background *B*: Bloggs has 420 tickets, and 580 other players have one ticket each.

The probability that Bloggs is the winner of the lottery is the same – .42 – against both backgrounds. However, native speakers judge (9) differently, depending on which background facts they are given. Against Background *A*, they robustly judge it false; against Background *B*, they robustly judge it true. Windschitl and Wells (1998) dubb this general phenomenon the *alternative outcomes effect*.

Windschitl and Wells conjecture that judgements in *B*-type cases are the ''product of an associative system that is primarily sensitive to pairwise comparisons rather than the more normatively appropriate comparisons between the focal outcome and the sum of all alternative outcomes'' (1413). Evidently they take it that the judgments in *B*-type cases are wrong (''normatively inappropriate''). But for the judgments to be wrong, (9) must be false; for (9) to be false, it must have a semantics that makes it false in a context where *B*-type information is presupposed. It appears that Windschitl and Wells assume that something a .5 absolute threshold semantics for *probably* is correct.

But a more plausible interpretation of the data is (not that speaker judgments are incorrect, but) that this view of the meaning of *probably* is incorrect. This sensitivity of *probably* to the contextually salient contrast alternatives substantially undermines the case for giving them a .5 threshold semantics, and makes it very plausible that they are relative, not absolute, adjectives. Here, I think, is a case where semantic research and psychological research can offer mutual illumination.

The obvious question to turn to now is, how is the relative contrast value for *probably* determined in context? Is it something like the *mean* of the probabilities of the live or salient alternatives? No, for consider Background *C*:

> Background *C*: Bloggs has 420 tickets, Smith has 520 tickets, and 60 other players have one ticket each.

Here the chance that Bloggs is the winner is appreciably greater than the mean of the salient alternatives, but we are not inclined to judge (9) true; we are inclined instead to say that Smith is probably the winner. Perhaps then we should say that the contrast value to ϕ is simply the alternative to ϕ with the greatest probability. But this too seems to be inadequate, at least without refinement. Suppose instead that the lottery is out of a million tickets; that Bloggs buys 300 tickets, and Smith buys 100; and that the other 999,600 tickets each go to 999,600 distinct players. Even though Bloggs is the most likely to win, and is well ahead of the next most likely alternative, native speakers will not enjoy affirming (9) against such a background. The big picture information that Bloggs's overall chances of winning are slim seems to swamp judgments.

We leave this problem open. In moving forward on it, it makes sense to look for an answer that will pattern generally with other relative adjectives. For discussion of several proposals concerning how the comparison value of a relative adjective is determined, see Kennedy (2007) and references cited therein. Additional insight may come from investigating what parallels there may be with the quantifier *most*. As with *probably*, it has been traditional to give *most* a ''more-than-half'' semantics. But that traditional semantics has come in for extended criticism recently by Hackl (2009), who presses evidence that *most* has comparative, ''relative superlative'' readings, as in 'John has read the most books, but he hasn't read more than half the books'.

Finally, we should ask: does the relativity of *probably* mean we must accept the failure of (V1)? That result is difficult to accept, for although speakers judge (9) true against a background like *B*, the sentence:

(10) # Bloggs is probably the winner of the lottery though it is more likely than not that he lost.

appears to be reliably terrible, regardless of the background information supplied.

An option less dramatic than abandoning (V1) suggests itself. When (9) is judged true against a background like *B*, the proposition that Bloggs is not the winner of the lottery is simply not one of the live alternatives; the question of its probability *per se* is not at issue. But in (10), this proposition is forced to become live, as it is the meaning of one of the clauses; and thereby judgements shift. To flesh this story out, it must be supposed that probability sentences are evaluated relative, not just to a probability space, but to a select class of propositions from the space (what we could call the *alternatives* set). It appears safe to assume that the alternatives set will usually constitute a partition of the sample space, and will not be closed under Boolean combinations.

7. *Scopal interactions*

A further semantic issue concerns the constraints on possible relations between probability operators and other scope-taking expressions. We have focused entirely on what we could call *de dicto* uses of probability operators. Are there not cases of quantifying into the scope of a probability operator? Consider first:

(11) Everyone probably lost the lottery.

The most natural, perhaps only, reading of this sentence is the one where *probably* takes wide scope with respect to *everyone*. This is surprising, as the narrow scope reading would make for intuitively more plausible truth-conditions (for most familiar lotteries, alas). This illustrates a case of what von Fintel and Iatridou (2003) call the *epistemic containment principle* (ECP), roughly the generalization that a quantifier cannot take scope over an epistemic modal in the same clause.[19] Still, as they note, one can force *de re* readings with the right multiclausal construction:

(12) There is somebody from New York who is likely to win the lottery. (von Fintel and Iatridou 2003: 190; building on Fox 2000)

Indefinite articles and tenses also make for interesting interactions, as recognized by Swanson (2006) (see also Swanson 2010, forthcoming). Clearly

(13) Bob is a likely choice for team captain.

is not equivalent to the claim that it's likely that Bob is a choice for team captain. Rather, its meaning is closer to something like

(14) There is a good chance that Bob will be chosen for team captain.

where the probability operator ('good chance') takes wide scope over a future tensed sentence. Similarly, if the headline reads

(15) Apple selects an unlikely design for their new phone.

once again the probability operator does not take wide scope. Rather, the meaning of (15) is closer to the superficially more complex:

(16) The design that Apple in fact selected is such that it was unlikely that it would be selected.

where the scopal order seems to be something like DEFINITE > PAST > UNLIKELY > WOULD.

Observe that this paraphrase eliminates the indefinite in (15), as does our paraphrase (14) of (13). By contrast, in

(17) Oskar Schindler was an unlikely hero.

the natural paraphrase is

(18) It was unlikely that Oskar Schindler would be a hero.

Here the indefinite remains, and the scopal order is PAST > UNLIKELY > WOULD > INDEFINITE. How exactly to give compositional semantics for sentences such as (13), (15) and (17) is an open problem. Given the evident complexity of the necessary paraphrases, it is tempting to question the assumption (as does Swanson (2006, ch. 2)) that these occurrences of *probably* and *likely* are true sentential operators, rather than the term level adjectives they superficially appear to be.

8. Embedding Potential

We close with some discussion of embedding potential of probability operators. The probability space semantics presented above is an extension of a classic relational modal semantics. According to this semantics, the state of information relevant to the interpretation of probability operators (and epistemic modals, and indicative conditionals) varies as a function of the evaluation world—just as, say, the knowledge state relevant to the interpretation of a knowledge operator varies as a function of the evaluation world in a normal epistemic logic. In such a logic, what worlds are epistemically accessible depends on what world one considers. If you shift the world of evaluation with an intensional operator, you potentially shift the set of epistemically accessible worlds. Now, if sentences like PROBABLY ϕ are likewise understood to characterize how things stand with respect to some body of information determined as a function of the evaluation world, we should expect them to behave, under intensional operators, in a manner broadly parallel to overt knowledge operators. But in fact, these sentences do not behave like knowledge ascriptions under under intensional operators. Consider embedding under verbs of imagination:

(19) John imagines that it is raining but that he doesn't know it is raining.

(20) # John imagines that it is raining but that it is not likely that it is raining.

There is nothing semantically defective about (19). According to it, John is entertaining the truth of a Moore-paradoxical proposition. (That is a slightly weird thing to be doing, of course, but it is perfectly doable, and there is nothing weird about the *sentence* describing what John is doing.) By contrast (20) is clearly marked—presumably because it is not at all clear what it even means to imagine both that it is raining and that it is not likely that it is raining. This contrast is not predicted on any of the semantic accounts we have considered; according to them we should expect the probability operator to behave like an attitude verb in this embedded context. So there is some important semantic asymmetry between attitude operators and epistemic modals/probability operators that is being missed on these accounts. See Yalcin (2007), where this worry is developed in detail. The solution recommended there, transposed to the probability space semantics considered above, is to relativize extensions, not to functions from worlds to epistemic probability spaces, but rather to epistemic probability spaces directly, and to adjust the semantics for attitude verbs, epistemic modals, and indicative conditionals accordingly. (This move may shed light also on why iterating probability operators tends to be odd (and why, when not odd, it is often vacuous):

(21) ? Probably, Sally is likely to be at the party.

Relevant here is the phenomenon of *modal concord*, wherein two modals tend to get read as one; see Geurts and Huitink (2006).)

An approach similar in spirit would be a dynamic implementation of the probability space semantics. In the dynamic semantics tradition (see Beaver (2001) for an overview), the meaning of a sentence is represented as a function (a 'context change potential' or 'update function') from a state of information to a state of information. On one well known version, a state of information is taken to be a set of possible worlds (or a set of possible world-sequence of individual pairs). But if states of information were represented as probability spaces, one might represent the meaning of a *probably* ϕ-clause as a 'test', one which maps a state of information to itself if it assigns to ϕ a high enough probability, and which otherwise maps to the null state. That would be a way of representing the idea that the clause expresses a condition on a possible state of information, one which parallels the classic dynamic treatments of epistemic modals (Veltman 1985, 1996). Alternatively, one might exploit the new options for dynamics which become available when information states are enriched probabilistically. For instance, one might represent the meaning of a *probably* ϕ-clause as a function which takes an information state i to the information state i' which makes ϕ likely, but which otherwise 'minimally changes' the probabilities assigned by i. Here, there is rich potential for semantics to benefit from work in formal epistemology, where various rigorous notions of 'minimal change' have been investigated (Halpern (2003) provides a summary of much of this work).

These matters are relevant when we move beyond semantics, to the question of what the informational content of probability claims is.[20] When one says that it is probability raining outside, what information exactly does one communicate? Is one here just *describing* one's state of information with sentence? Or is it rather that one is *expressing* one's state of information by saying it is probably raining, without literally saying that one is in such a state of information? (Cf. Gibbard (1990) on normative discourse.) If the sort of view advanced by Yalcin (2007, forthcoming) is right, and probability sentences semantically express conditions on states of information which are not reducible to conditions on possible worlds (ways the world might be, or truth-conditions in the usual sense), then the latter view—which might be called *content probabilism* or *credal expressivism*, depending on how it is developed—becomes a live possibility. But again, this issue goes beyond semantics.

Acknowledgements

Earlier versions of this material were presented at MIT, Sinn and Bedeutung 11, and at a graduate seminar at the University of Massachusetts, Amherst, in 2006; at Arché in 2008; at my NYU graduate seminar in 2009; and at Berkeley linguistics in 2009. I am indebted to audiences at all these places, and in particular to conversations with, and/or suggestions from, Kai von Fintel, Michael Glanzberg, Valentine Hacquard, Angelika Kratzer, Daniel Lassiter, Dilip Ninan, Paul Portner, Daniel Rothschild, and Bob Stalnaker. Special thanks to Eric Swanson for detailed comments on the penultimate draft.

Short Biography

Seth Yalcin is an assistant professor of philosophy at the University of California, Berkeley, where he is also member of the Group in Logic and Methodology of Science. He works in the philosophy of language and related areas.

Notes

* Correspondence: Department of Philosophy, University of California, Berkeley, CA 94720, USA. Email: yalcin@berkeley.edu

[1] For instance, it is common in that literature to begin with semantics for constructions like *Bob is taller than Steve* and work back from there to a semantics for *Bob is tall*, treating the latter as expressing a covert comparative, one which compares Bob's height to something like a contextually salient standard. See Kennedy (2007) for a recent discussion.

[2] Since these words interact with all the morphology usual to gradable adjectives (*more, less, -er, as, than, too, enough, quite, very*, etc.), and since we want to give as uniform a semantics for all this morphology as possible, we should expect the semantics of our probability operators to eventually be unified with that of gradable predicates generally.

[3] Technical features of the model assumed by Kratzer are discussed independently by Halpern (1997, 2003: 2.7). Both Kratzer and Halpern credit Lewis's semantics for counterfactuals (Lewis 1973, 1981) for essential ideas. (Note that the semantics for probability operators given by Kratzer (1991) differs from that given by Kratzer (1981). I focus on her most recent account.)

[4] To discuss valid inference, we of course must assume some notion of consequence. The natural way to think of consequence on Kratzer's semantics (and the way she in fact thinks of it) is simply in terms of preservation of truth at a world, given a fixed context. That is the understanding assumed here.

[5] Unless otherwise noted, take my schematic letters to be replaceable by present-tensed 'unhedged' sentences, sentences which are themselves not probabilistic or epistemically modalized.

[6] Kratzer's approach to *if*-clauses builds on Lewis (1975). On the usual development of this analysis, *if*-clauses are taken to semantically combine with modal operators; they are modal modifiers. But see also Kratzer (1991), where *if*-clauses are treated as modal-base-shifting operators.

[7] 'Positive form' here is a reference to the fact that *probable* is a positive form adjective. It seems positive form adjectives in general license this sort of pattern. (For example: if John is at least as tall as Niko, and Niko is tall, it follows that John is tall.)

[8] Consider a set of epistemically accessible worlds 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 $w' \geq w$, $w' \in p$), 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 assumption of totality is made entirely for convenience here. Add if you wish a second, disconnected infinite chain of worlds, one which forever alternates as we ascend between p and $\neg p$-worlds, and between q and

$\neg q$-worlds. (Indeed add any number of such chains, branching in myriad ways if you please.) Now we no longer have a total order, but we still have a case where, for Kratzer, p is probable, q is as likely as p, but q is not probable.

This counterexample assumes 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 only for convenience) a totally ordered set of worlds, and (making the limit assumption) now consider the set M of worlds which are maximal according to the order. Let p be true throughout the M-worlds, but q be true only throughout a proper subset of M-worlds. 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 between worlds be allowed?

[9] Hamblin (1959) appears to have been the first to formalize the notion 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.

[10] We are spotting Hamblin some modern ideas to get his account on the table.

[11] More precisely: Kratzer (1991) does not validate this pattern. The account in Kratzer (1981) *does* validate this pattern.

[12] 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, since in a normal modal logic, it is a theorem that $(\Box\phi \wedge \Box\psi) \supset \Box(\phi \wedge \psi)$.

[13] Hamblin offers the following example in support of the idea that Conjunctivitis is valid: 'The murder-weapon was probably a stiletto and it probably pierced the heart; hence death was probably immediate' (Hamblin 1959: 238). But the *it* in the second conjunct, despite being under a new probability operator, is naturally understood as anaphoric on 'a stiletto', which leads the interpreter to understand one possibility, not two separate possibilities, being described as probable—a clear case of modal subordination. (Eventually, of course, an account of probability operators must be integrated with a plausible account of modal subordination. I regret I cannot explore the options here.)

[14] More precisely: $E^{\phi}_{e(w)} =_{\text{DEF}} \{w' \in E_{e(w)} : [[\phi]]^{w',e} = 1\}$.

[15] More precisely: $Pr^{\phi}_{e(w)}(p) =_{\text{DEF}} Pr_{e(w)}(p|\{w' : [[\phi]]^{w',e(w)} = 1\})$. We assume the standard ratio definition of conditional probability ($Pr(p\,|\,q) =_{\text{def}} Pr(p \cap q)/Pr(q)$).

[16] When (as here) *if*-clauses are treated as operators rather than modal modifiers, the question of the relative scope of the clause and a modal operator can arise. This is an advantage if the predicted scope ambiguities exist, a disadvantage if they do not. The debate about which of these treatments of *if*-clauses is preferable is beyond our scope; but see Rothschild (forthcoming) for an interesting recent discussion of some of the relevant issues.

[17] One must, of course, eventually generalize the story beyond this restricted fragment of conditional talk I am focused on—to probability operator-containing antecedents and to counterfactual antecedents, for instance. (Counterfactual *if*-clauses presumably induce a different kind of shift on e; imaging on the antecedent, perhaps (Lewis 1976.)

[18] Note that this proposal already makes probability operators exceptional, since paradigm absolute adjectives express relations to minimum or maximal scale values, not values in between.

[19] The fact that ordinary epistemic modals and probability operators tend to obey this constraint supplies additional, syntactic evidence for classifying them together.

[20] I take it there is a gap between formal semantics and the theory of content; 'compositional semantic value' and 'content' correspond to different theoretical roles. See Lewis (1980), Yalcin (2009).

Works Cited

Beaver, David. *Presupposition and Assertion in Dynamic Semantics*. Studies in Logic, Language and Information. Stanford, CA: CSLI Publications, 2001.

Cohen, Ariel. 'Probability in Semantics.' *Language and Linguistics Compass* 3.1 (2009): 265–81.

von Fintel, Kai and Sabine Iatridou. 'Epistemic Containment.' *Linguistic Inquiry* 34.2 (2003): 173–98.

—— and Anthony S. Gillies. 'An Opinionated Guide to Epistemic Modality.' *Oxford Studies in Epistemology* 2 (2007): 32–62.

Fox, Danny. *Economy and Semantic Interpretation*. Cambridge, MA: The MIT Press, 2000.

Geurts, Bart and Janneke Huitink. 'Modal Concord.' *Concord Phenomena and the Syntax Semantics Interface*. Ed. Paul Dekker and Hedde Zeijlstra. Malaga: ESSLLI, 2006. 15–20.

Gibbard, Allan. *Wise Choices, Apt Feelings*. Cambridge: Harvard UP, 1990.

Hackl, M. 'On the Grammar and Processing of Proportional Quantifiers: Most Versus More Than Half.' *Natural Language Semantics* 17.1 (2009): 63–98.

Hájek, Alan. 'What Conditional Probability Could not be.' *Synthese* 137 (2003): 273–323.

——. 'Interpretations of Probability.' *Stanford Encyclopedia of Philosophy*. 2009. 1 June 2010 <http://plato.stanford.edu/entries/probability-interpret/>.

Halpern, Joseph Y. 'Defining Relative Likelihood in Partially-Ordered Preferential Structures.' *Journal of Artificial Intelligence Research* 7 (1997): 1–24.

——. *Reasoning about Uncertainty*. Cambridge, MA: MIT Press, 2003.

Hamblin, Charles L. 'The Modal "Probably".' *Mind* 68.270 (1959): 234–40.

Kennedy, Chris. 'Vagueness and Grammar: The Semantics of Relative and Absolute Gradable Adjectives.' *Linguistics and Philosophy* 30.1 (2007): 1–45.

Kolodny, Niko and John MacFarlane. 'Ifs and Oughts.' *Journal of Philosophy* CVII.3 (2010): 115–43.

Kratzer, Angelika. 'The Notional Category of Modality.' *Words, Worlds, and Contexts, New Approaches to Word Semantics*. Ed. H.-J. Eikmeyer and H. Rieser. Berlin: Walter de Gruyter, 1981. 38–74.

——. 'Conditionals.' *Papers from the Parasession on Pragmatics and Grammatical Theory*. Ed. A. Farley, P. Farley and K.-E. McCullough. Chicago, IL: Chicago Linguistics Society, 1986. 1–15.

——. 'Modality.' *Semantics: An International Handbook of Contemporary Research*. Ed. A. von Stechow and D. Wunderlich. Berlin: de Gruyter, 1991. 639–50.

Kyburg, Henry. 'Conjuctivitis.' *Induction Acceptance and Rational Belief*. Ed. M. Swain. Dordrecht: Reidel, 1970. 55–82.

Lewis, David K. *Counterfactuals*. Malden: Blackwell, 1973.

——. 'Adverbs of Quantification.' *Formal Semantics of Natural Language*, Ed. E. L. Keenan. Cambridge: Cambridge UP, 1975. 3–15.

——. 'Probabilities of Conditionals and Conditional Probabilities.' *Philosophical Review* 85 (1976): 297–315.

——. 'Index Context and Content.' *Philosophy and Grammar*. Ed. S. Kanger and S. Ohman. Holland: D. Reidel, 1980. 79–100.

——. 'Ordering Semantics and Premise Semantics for Counterfactuals.' *Journal of Philosophical Logic* 10 (1981): 217–34.

Mellor, D. H. *Probability: A Philosophical Introduction*. New York, NY: Routledge, 2005.

Portner, Paul. *Modality*. Oxford: Oxford UP, 2009.

Roberts, Craige. 'Modal Subordination and Pronominal Anaphora in Discourse.' *Linguistics and Philosophy* 12.6 (1989): 683–721.

Rothschild, Daniel. 'Do Indicative Conditionals Express Propositions?' *Nous* (forthcoming).

Shackle, George L. S. 'On the Meaning and Measure of Uncertainty.' *Metroeconomica* 5.3 (1953): 97–115.

Shafer, G. *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton UP, 1976.

Swanson, Eric. *Interactions with Context*. PhD thesis. Cambridge MA: Massachusetts Institute of Technology, 2006.

——. 'Modality in Language.' *Philosophy Compass* 3.6 (2008): 1193–207.

——. 'On Scope Relations Between Quantifiers and Epistemic Modals.' *Journal of Semantics* (2010): 1–12.

——. 'How not to Theorize About the Language of Subjective Uncertainty.' *Epistemic Modality*. Ed. Andy Egan and Brian Weatherson. Oxford: Oxford UP, forthcoming.

Tversky, A. and D. J. Koehler. 'Support Theory: A Nonextensional Representation of Subjective Probability.' *Psychological Review* 101.4 (1994): 547–67.

Veltman, Frank. *Logics for Conditionals*. PhD thesis. University of Amsterdam, 1985.

——. 'Defaults in Update Semantics.' *Journal of Philosophical Logic* 25.3 (1996): 221–61.

Williamson, Timothy. *Knowledge and its Limits*. Oxford: Oxford UP, 2000.

Windschitl, Paul D. and Gary L. Wells. 'The Alternative-Outcomes Effect.' *Journal of Personality and Social Psychology* 75.6 (1998): 1411–23.

Yalcin, Seth. 'Epistemic Modals.' *Mind* 116.464 (2007): 983–1026.

——. 'More on Epistemic Modals.' *Mind* 118.471 (2009): 785.

——. 'Nonfactualism About Epistemic Modality.' *Epistemic Modality*. Ed. Andy Egan and Brian Weatherson. Oxford: Oxford UP, forthcoming.

Zadeh, Lofti A. 'Fuzzy Sets as a Basis for a Theory of Possibility.' *Fuzzy Sets and Systems* 1 (1978): 2–28.