Epistemic comparison, models of uncertainty, and the disjunction puzzle

Daniel Lassiter*
Stanford University

Pre-final version of an article in press at *Journal of Semantics*.
Please consult the final version at http://jos.oxfordjournals.org/.

Abstract The best-known theory of modality in linguistics (Kratzer 1991, 2012) uses a binary relation on worlds to state truth-conditions for sentences with epistemic auxiliaries, and lifts this order to an order on propositions to state meanings for epistemic comparatives and equatives with *likely* and *probable*. It has recently been observed that this theory makes incorrect predictions about the truth-conditions of epistemic comparative and equative sentences. I analyze the source of this problem in Kratzer's and several related theories, and consider several modifications suggested in the recent literature. All of the theories available either fail to resolve the problem, or generate incorrect predictions about the relationship between *must* and *likely* when combined with Kratzer's semantics for the auxiliaries.

Various alternative models of uncertainty are available which can be used as a semantics for *likely* and do not encounter the problem with disjunction. In particular, qualitative and numerical probability have the necessary features, as do several weaker approaches which include probability as a special case. Data involving entailments and degree modification are argued to provide a reason for preferring the probabilistic approach. I then consider several methods of integrating this approach with Kratzer's semantics for the epistemic auxiliaries, showing that all of them wrongly invalidate the inference from *must* to *(much) more likely than not*. I conclude by stating three alternative semantics for the auxiliaries which do better, one quantificational and two probabilistic. The choice between them will depend on the resolution of three open problems involving the meanings of the auxiliaries which are detailed in the concluding section.

Keywords: Epistemic modality, ordering semantics, premise semantics, modal comparison, probability, modal auxiliaries

The theory of modality associated with Kratzer (1981, 1991, 2012) has been extremely influential in linguistic semantics, and deservedly so. It has numerous advantages over the classic treatment of modals as quantifiers over unstructured sets of accessible worlds stemming from modal logic, for example in its elegant treatment of the interaction between modals and conditionals, and its robustness in the face of conflicting information and obligations. One of the most important empirical virtues of this theory lies in its ability to treat intermediate and comparative modalities.

* Warm thanks to the many people who gave me helpful feedback on versions of this paper: Thomas Icard, Wes Holliday, Seth Yalcin, Sven Lauer, Cleo Condoravdi, Chris Barker, Annie Zaenen, Lauri Karttunen, and in particular Robert van Rooij and the *JoS* reviewers, whose critical comments led to many improvements.

While the simpler classic approach gives us plausible analyses of "extreme" modals such as *must*, *necessary*, *can*, and *might*, it founders on such basic modal concepts as intermediate *good*, *likely*, and *probable* as well as comparative modalities such as *as likely as*, *better than* and *more desirable than*. Kratzer's theory, on the other hand, is able to assign non-trivial truth-conditions for sentences containing intermediate and comparative modals by means of a relation of *comparative possibility* on propositions. In its ability to analyze this important class of modals, Kratzer's theory is a clear improvement on the classic approach.

Unfortunately, the critical notion of "Comparative Possibility" in Kratzer 1991 makes unwelcome predictions about the truth-conditions of epistemic comparatives: this theory predicts that (1a) and (1b) should together entail (1c). I call this the DISJUNCTION PUZZLE.

(1) a. ϕ is as likely as ψ.
b. ϕ is as likely as χ.
c. $\therefore$ ϕ is as likely as $(\psi \vee \chi)$.

Using this validity we can prove some quite counter-intuitive consequences, for example the prediction that (2a) entails (2b).

(2) a. Sam is as likely to win the lottery as anyone else is.
b. $\therefore$ Sam is as likely to win the lottery as he is not to win it.

Suppose that (2) were valid. Then, if Sam buys as many lottery tickets as the customer who has bought the most, he is assured of a serious chance of great wealth in the near future. But of course this inference isn't valid: in a typical lottery, if each customer buys 1 or 2 tickets each, then (sad to say) each customer is overwhelmingly less likely to win than not to win.

Section 1 of this paper is devoted to explicating the disjunction puzzle and its ramifications for the standard theory and related approaches, notably a modification of Lewis's (1973) proposal involving counterfactuals and deontic comparatives. The problematic feature of these theories is that the likelihood of a disjunction is equal to the maximum of the likelihood of the disjuncts in each of these theories. While the picture in Kratzer's (1991) approach is complicated by the weaker structure of the underlying binary order, Lassiter (2010a) and Yalcin (2010) independently observed that this theory also predicts (1) valid. I give a new proof showing that the revised definition of Comparative Possibility proposed by Kratzer (2012), intended to resolve a special case of this puzzle noted by Yalcin (2010), leaves the core of the problem untouched.

A new method of generating a likelihood ordering from an ordering on worlds due to Holliday & Icard (2013a) does make the right predictions about disjunction (and appears to behave well when combined with one of the proposals for *must* and *might* discussed in §4). However, the revision is little help for the overall project of rescuing Kratzer's framework: when combined with her semantics for the auxiliaries, it renders invalid the clearly valid inference in (3).

(3) a. ϕ must be the case.
b. ϕ is more likely than $\neg\phi$.

Section 2 discusses several ways to avoid these problems. A straightforward way to avoid the issue with disjunction is to adopt a semantics for *likely* built around qualitative or quantitative probability, which rely crucially on some form of *additivity*. This approach can be stated using standard tools of degree semantics. Several weaker degree theories are available which also resolve the puzzle. The first, based on *symmetric fuzzy measures*, is shown to have other empirical problems. The second, based on *qualitatively additive measures*, is discussed in detail by Holliday & Icard (2013a), who show that it shares numerous desirable features of the probabilistic account. These authors correctly point out that it is difficult to motivate the stronger semantics based on entailment data alone. However, I argue that the acceptability of ratio modifiers such as *exactly twice as ... as* with *likely* provides reason to prefer the probabilistic semantics, as long as previous work on modification is correct in claiming that these modifiers are only acceptable with adjectives on additive scales (such as *heavy* and *tall*).

Section 3 turns to implications for the epistemic auxiliaries. If we wish to maintain Kratzer's (1991) semantics for *must* and *might* after adopting a probabilistic semantics for *likely*, bridging rules are needed to enforce logical relations between these expressions. I examine three salient possibilities, showing that they all fail because the intuitively valid inference pattern in (3) comes out as invalid. This does not prove conclusively that no such combination is possible, but it does lay out clearly the empirical and formal challenge which will have to be met if Kratzer's framework is to be viable.

Section 4 considers three alternative semantics for the auxiliaries: one in which they are quantifiers over *all* epistemically possible worlds, and two in which they constrain probability assignments directly. Several empirical issues which could in principle distinguish between these similar accounts. After considering and discarding one based on continuous sample spaces, I discuss three open issues whose resolution should allow us to choose: the "weakness" of *must*, the question of whether ϕ *is more likely than* ψ entails *might* ϕ, and the proper representation of awareness in a probabilistic semantics for the auxiliaries.

1 Epistemic comparison and disjunction

1.1 Empirical preliminaries

Imagine a lottery with 1 million tickets. Three siblings — Sam, Mary, and Sue — buy two tickets each. No one else buys more than two tickets. The lottery is fair, and only one ticket will be chosen as the winner. (4) is clearly true here, since Sam has as many tickets as anyone else does.

(4) Sam is as likely to win the lottery as anyone else is.

What happens when the proposition in the comparative complement is a disjunction? Consider (5):

(5) Sam is as likely to win the lottery as Mary or Sue is.

(5) seems to be true, but this is not really relevant to the interaction of *or* with epistemic comparatives. (5) is true for the same reason that (6) is, assuming that Sam's height is as great as Mary's and as great as Sue's.

(6) Sam is as tall as Mary or Sue is.

For extraneous reasons (on which see e.g. Schwarzschild & Wilkinson 2002; Lassiter 2010b) disjunctions in the complement of comparatives and equatives can be interpreted as if they were wide-scope conjunctions, so that (6) is equivalent to (7).

(7) Sam is as tall as Mary is, and Sam is as tall as Sue is.

Similarly, (5) is true in the lottery scenario described above because it is read as meaning "Sam is as likely to win as Mary is, and Sam is as likely to win as Sue is". To control in part for this alternative interpretation of *or*, consider instead (8) and (10):

(8) It is as likely that Sam will win as it is that Mary or Sue will win.

(9) It is as likely that Sam will win as it is that one of his sisters will win.

On one prominent reading, (8) is false: Mary and Sue have four tickets together, and it is more likely that one of them will win than it is that Sam — who holds only two tickets — will win. The relevant reading is equivalent to (9), which has no overt disjunction and is unambiguously false in the scenario at hand. I assume that this reading is adequately captured by (10), with $\succcurlyeq_{likely}$ representing a transitive and reflexive relation of comparative likelihood between propositions.

(10) $\mathbf{win}(\mathbf{Sam}) \succcurlyeq_{likely} [\mathbf{win}(\mathbf{Mary}) \vee \mathbf{win}(\mathbf{Sue})]$

The fact that (8)/(9)/(10) is false here indicates that the inference pattern in (11) is not valid:

(11)
a. $\phi \succcurlyeq_{likely} \psi$.
b. $\phi \succcurlyeq_{likely} \chi$.
c. $\therefore \ \phi \succcurlyeq_{likely} (\psi \vee \chi)$.

Now consider another handful of people: Tom, Alice, Bob, Hank, and Murray, who also bought two tickets each. It is even more clear that (11) is false on the relevant reading:

(12) It is as likely that Sam will win as it is that Mary, Sue, Tom, Alice, Bob, Hank, or Murray will.

If the inference pattern in (11) were valid, (12) would be a valid inference as well, since we could feed the conclusion of an application of (11) back into itself as a premise, starting with two individuals and adding one more in each iteration. Indeed, if we had an exhaustive list $\{x_1, x_2, ... x_n\}$ of the n people other than Sam who bought the other $999{,}998$ lottery tickets (with no more than two each), we would eventually be able to prove:

(13)
a. $\mathbf{win}(\mathbf{Sam}) \succcurlyeq_{likely} \mathbf{win}(x_1)$
b. $\mathbf{win}(\mathbf{Sam}) \succcurlyeq_{likely} \mathbf{win}(x_2)$
c. $\therefore \ \mathbf{win}(\mathbf{Sam}) \succcurlyeq_{likely} [\mathbf{win}(x_1) \vee \mathbf{win}[(x_2)]$
d. $\mathbf{win}(\mathbf{Sam}) \succcurlyeq_{likely} \mathbf{win}(x_3)$
e. $\therefore \ \mathbf{win}(\mathbf{Sam}) \succcurlyeq_{likely} [\mathbf{win}(x_1) \vee \mathbf{win}(x_2) \vee \mathbf{win}(\mathbf{x}_3)]$
f. ...
g. $\therefore \ \mathbf{win}(\mathbf{Sam}) \succcurlyeq_{likely} [\mathbf{win}(x_1) \vee \mathbf{win}(x_2) \vee \mathbf{win}(x_3) \vee \cdots \vee \mathbf{win}(x_n)]$

Since the lottery has one and only one winner, the proposition on the right side of (13g) — that someone other than Sam wins — is equivalent to the proposition that Sam does not win.

By this route, if (11) were valid we could prove:

(14) $\textbf{win}(\textbf{Sam}) \succcurlyeq_{likely} \neg\textbf{win}(\textbf{Sam})$

But (14) is, as noted above, clearly false here: Sam, having only two of the 1,000,000 tickets sold in a fair lottery, is overwhelmingly more likely not to win than he is to win. As a result, a theory which predicts that (11) is valid would make clearly incorrect predictions about the entailments of comparative likelihood statements, and so does not correctly capture their truth-conditions.

A few observations before we proceed to the theoretical discussion. First, since there is only one winner in the lottery at hand, the propositions whose likelihood we are comparing happen to be mutually exclusive: for any x and y distinct from Sam, $x \neq y$, if x wins then y does not, and if Sam wins then neither x nor y does. The inference pattern discussed in (11) does not require this, but it also does not exclude it. This point will be important in §1.4 below.

Second, the use of quantifiable objects such as numbers of lottery tickets in this example is not crucial to make the empirical point. The same reasoning holds in any domain where there are three or more propositions under consideration, and intuitions are particularly clear on this point if the propositions are mutually exclusive. Sports commentators forecasting the outcome of a large tournament, for example, would in most cases happily agree that, if team A is clearly the best team in the tournament — i.e., A is pairwise clearly more likely to win than each of its competitors — A may still be less likely to win the tournament than they are not to win, since they may have many opportunities to lose against the odds.[1]

1.2 Roots of the disjunction puzzle

Before considering the treatment of inferences like (1) in Kratzer's theory, it may be helpful to consider briefly some earlier and related approaches. All of these accounts share the problematic predictions involving disjunction under consideration.

Lewis (1973) gives several equivalent versions of his semantics for counterfactuals, one of which is framed in terms of a notion of "comparative possibility" (see his §2.5). This system is logically very similar to Kratzer's semantics for modals. While Lewis does not consider the possibility of treating his theory as a semantics for epistemic comparatives, he does discuss an application to deontic comparatives (Lewis 1973: §5.1), and Halpern (1997, 2003) discusses explicitly the logic which results from taking Lewis' semantics for counterfactuals (without the assumption of centering) as a theory of *as likely as* and related notions.[2] The fact that such a semantics for epistemic comparatives encounters the disjunction puzzle is built directly into the axiom system.

1 For example, as of June 20, 2013 the sports betting website Bovada had chosen the University of Kentucky as the most likely victor in the 2014 NCAA Division 1 Men's Basketball Tournament. Bovada gave odds of 4:1 on Kentucky's victory, thus agreeing to pay anyone who bets on Kentucky four times what they bet if Kentucky does win in 2014. This offer, coming from a website whose business is to make money from sports betting, strongly suggests that (as of that date) the team that Bovada considered most likely to win was nevertheless thought to be substantially more likely *not* to win — indeed, at least four times as likely.

2 Halpern's discussion is inspired by Lewis (1973), and he does not appear to be aware of Kratzer's closely related work. Halpern concentrates on a version of the theory without the connectedness assumption and a rule relating world- and proposition-orderings equivalent to (15). This semantics admits the same class of models as Kratzer's theory of epistemic modality, for the same reason that Kratzer's premise semantics for counterfactuals is equivalent to Lewis' ordering semantics without the assumption of connectedness (proved in Lewis 1981).

A minimal modification of Lewis' version of comparative possibility for the epistemic case might look like this (dropping the centering axiom, "all and only truths are maximally possible", which would be inappropriate for epistemic language). We assume a transitive, connected, reflexive ordering on propositions $\succcurlyeq_{likely}$ which also satisfies two further axioms. Lewis' axiom (4) requires, for any set of propositions A, that $\phi \succcurlyeq_{likely} \bigcup A$ if and only if $\phi \succcurlyeq_{likely} \psi$ for each $\psi \in A$. His axiom (5) requires that, if $\{w\} \succcurlyeq_{likely} \psi$ for every ψ in a set of propositions A, then $\{w\} \succcurlyeq_{likely} \bigcup A$.[3]

As Lewis points out, axiom (4) entails that the disjunction of a finite set of propositions A is exactly as likely as the most likely member of A. Axiom (4) thus has the effect of ensuring that the problematic inference in (1) is satisfied (see Halpern 1997 for additional discussion).

Lewis also demonstrates that orderings of propositions that satisfy his comparative possibility axioms can be put into 1-to-1 correspondence with a reflexive, transitive, connected ordering $\succcurlyeq$ on worlds. This correspondence is useful not only because it clarifies how comparative possibility relates to Lewis' other (sphere- and comparative similarity-based) semantics, but also because it allows us to see how we could derive comparative possibility from a more basic ordering on worlds, as Kratzer (1981, 1991, 2012) and Halpern (1997) do. If we choose to set things up this way, then an ordering on propositions which satisfies the axioms described above can be generated by a simple rule (Lewis 1973: 56, 49):

(15) $\phi \succcurlyeq_{likely} \psi$ if and only if, for every world w in ψ, there is a world w' in ϕ such that $w' \succcurlyeq w$.

This is also Kratzer's method of deriving an ordering on propositions from an ordering on worlds, as we will see below.

Starting from the perspective in which orderings on worlds are basic, then, we can identify the root of the disjunction puzzle in the modified Lewisian semantics: if there is even a single ϕ-world which is ranked as high by $\succcurlyeq$ as all ψ-worlds, then ϕ is at least as likely as ψ. Informally, we can gloss this feature as follows: the position of a proposition in the likelihood ordering is determined exclusively by the position of the highest-ranked world(s) in the proposition, and all information about lower-ranked worlds is ignored. (Note however that this gloss is technically correct only in models which are finite or satisfy the limit assumption.) In the finite case, then, the likelihood of a disjunction is equal to the likelihood of the most likely disjunct: the likelihood that someone other than Sam will win the lottery is equal to the likelihood that x will win the lottery, where x is an individual other than Sam who bought as many lottery tickets as anyone else. This may well be a reasonable prediction for counterfactuals and deontic comparisons (but see e.g. Goble 1996; Lassiter 2011, 2014a for criticism in the deontic case). However, as we saw in §1.1, the prediction is clearly incorrect for epistemic comparatives.

A number of other theories that have been applied to graded and comparative epistemic language share the property that the likelihood of a (finite) disjunction is equal to the likelihood of the most likely disjunct or disjuncts. For example, Hamblin's (1959) metrical semantics for *probable* has this feature, as discussed by Yalcin (2010). In a semantics which mimics classical fuzzy logic (Zadeh 1965, 1978), (11) would be valid because the likelihood of $\phi \vee \psi$ is simply defined to be the likelihood of ϕ or that of ψ, whichever is greater.

3 Note that my notation differs from Lewis' and Kratzer's in reversing the direction of the relations. I do this because, with epistemic comparatives, $\succ$ and $\succcurlyeq$ correspond intuitively to "more than" and "at least as much as" rather than "less than", etc.

1.3 The disjunction puzzle in Kratzer 1991

In the theory of Kratzer (1991), sentences containing modals of all stripes (not just epistemic modals) are assigned truth-conditions by the interaction of three factors:

- The *modal base* $\mathbf{f}$, a function from worlds to sets of propositions;
- The *ordering source* $\mathbf{g}$, a function from worlds to sets of propositions;
- The lexical semantics of the modal word, which determines what use it makes of the modal base and the ordering source.

In many contexts, $\mathbf{f}$ and $\mathbf{g}$ are left as free parameters to be determined pragmatically, possibly subject to constraints from the lexical semantics of particular expressions. Different modal "flavors" are modeled by varying the particular choice of $\mathbf{f}$ and $\mathbf{g}$ in a context.

As in classic modal logic, the truth-conditions of sentences containing strong modals (*must, should, necessarily, obligatorily, ...*) and weak modals (*may, might, possibly, permissibly, ...*) are given in terms of universal and existential quantification respectively. For these items, the crucial new feature introduced by Kratzer is that the set of worlds quantified over is not simply given in advance by an accessibility relation, but determined in a somewhat complex fashion by the interaction of $\mathbf{f}$ and $\mathbf{g}$. First, instead of taking an ordering on worlds as primitive as we did in the modified Lewisian semantics discussed above, we derive an ordering on worlds from the sets of propositions $\mathbf{f}(w)$ and $\mathbf{g}(w)$ as follows:

(16) $u \succcurlyeq_{\mathbf{g}(w)} v$ if and only if $\{p \mid p \in \mathbf{g}(w) \wedge u \in p\} \supseteq \{p \mid p \in \mathbf{g}(w) \wedge v \in p\}$ $\qquad (u, v \in \bigcap \mathbf{f}(w))$

That is, for worlds u and v compatible with all of the propositions in the modal base, u is at least as good a world as v — $u \succcurlyeq_{\mathbf{g}(w)} v$ — if and only if u satisfies every proposition in the ordering source that v does, and possibly more (Kratzer 1991: 644).

Must(ϕ) is defined as in (17), and *might*(ϕ) as its dual $\neg$*must*$(\neg\phi)$.

(17) $[\![\textit{must } \phi]\!]^{\mathcal{M},w}_{\mathbf{f},\mathbf{g}} = 1$ iff $\forall u \exists v[v \succcurlyeq_{\mathbf{g}(w)} u \wedge \forall z : z \succcurlyeq_{\mathbf{g}(w)} v \rightarrow z \in \phi]$ $\qquad (u, v, z \in \bigcap \mathbf{f}(w))$

The complexity of this definition is due to a desire to avoid making the limit assumption (Kratzer 1991: 644). When this assumption is satisfied, and in models in which $\mathbf{g}(w)$ is consistent or finite, the definition simplifies to the condition that ϕ is true in all of the worlds in $\bigcap \mathbf{f}(w)$ that are maximal in $\succcurlyeq_{\mathbf{g}(w)}$.

(18) $\mathbf{BEST}(\mathbf{f}, \mathbf{g}, w) =_{df} \{v \mid v \in \bigcap \mathbf{f}(w) \wedge \neg\exists v' \in \bigcap \mathbf{f}(w) : v' \succ_{\mathbf{g}(w)} v\}$

(19) a. $[\![\textit{must } \phi]\!]^{\mathcal{M},w}_{\mathbf{f},\mathbf{g}} = 1$ iff $\forall u : u \in \mathbf{BEST}(\mathbf{f}, \mathbf{g}, w) \rightarrow u \in \phi$.

b. $[\![\textit{might } \phi]\!]^{\mathcal{M},w}_{\mathbf{f},\mathbf{g}} = 1$ iff $\exists u : u \in \mathbf{BEST}(\mathbf{f}, \mathbf{g}, w) \wedge u \in \phi$.

We will not make the limit assumption here, but we will occasionally make use of models in which there happen to be "best" worlds in order to show that certain inferences are not valid.

The truth-conditions of *probably* and *more likely than* are given in terms of a binary relation on propositions which is derived from the binary relation on worlds $\succcurlyeq_{\mathbf{g}(w)}$ using the same rule that Lewis (1973: §2.5) uses to lift an ordering on worlds to an ordering on propositions (cf. (15)).

(20) COMPARATIVE POSSIBILITY: ϕ is as good a possibility as ψ (relative to $\mathbf{f}$, $\mathbf{g}$, and w) if and only if, for every ψ-world $u \in \bigcap \mathbf{f}(w)$, there is a ϕ-world $v \in \bigcap \mathbf{f}(w)$ such that $v \succcurlyeq_{\mathbf{g}(w)} u$. (Kratzer 1991: 644)

We define derived "better possibility" and "exactly as good a possibility as" in the usual way:

(21) a. ϕ is a better possibility than ψ if and only if ϕ is a good a possibility as ψ, and ψ is not as good a possibility as ϕ.

b. ϕ is exactly as good a possibility as ψ if and only if ϕ is a good a possibility as ψ, and ψ is as good a possibility as ϕ.

Kratzer (1991: 644-5) adds the following definitions for *probably* and *more likely than*:

(22) a. $[\![\textit{probably } \phi]\!]^{\mathcal{M},w}_{\mathbf{f},\mathbf{g}} = 1$ if and only if ϕ is a better possibility than $\neg\phi$.

b. $[\![\phi \textit{ is more likely than } \psi]\!]^{\mathcal{M},w}_{\mathbf{f},\mathbf{g}} = 1$ if and only if ϕ is a better possibility than ψ.

Since we will be making frequent use of the relation "is as good a possibility as", I define an abbreviation $\succcurlyeq^o$, leaving relativization to $\mathbf{f}$ and $\mathbf{g}$ implicit except where there is potential for confusion. (The ^{o} is for "old", since this relation will shortly be replaced with a new order $\succcurlyeq^n$.)

(23) $\phi \succcurlyeq^o \psi$ iff$_{df}$ $\forall u \in \psi\ \exists v \in \phi : v \succcurlyeq_{\mathbf{g}(w)} u$, where u, v are restricted to $\bigcap \mathbf{f}(w)$.

$\succ^o$ represents "is a better possibility than", and $\approx^o$ represents "is exactly as good a possibility as". This is Kratzer's (1991) proposal for the truth-conditions of epistemic comparative and equative sentences, then:

(24) a. $[\![\phi \textit{ is more likely than } \psi]\!]^{\mathcal{M},w}_{\mathbf{f},\mathbf{g}} = 1$ if and only if $\phi \succ^o \psi$.

b. $[\![\phi \textit{ is as likely as } \psi]\!]^{\mathcal{M},w}_{\mathbf{f},\mathbf{g}} = 1$ if and only if $\phi \succcurlyeq^o \psi$.

However, Lassiter (2010a) and Yalcin (2010) independently observed that the argument in (11), restated in these terms, is valid (see the Appendix, §6.1 for a proof).

(25) a. $\phi \succcurlyeq^o \psi$

b. $\phi \succcurlyeq^o \chi$

c. $\therefore\ \ \phi \succcurlyeq^o (\psi \vee \chi)$

The problem is essentially the same as in the modified Lewisian semantics: the rule lifting world-orderings to proposition-orderings is defined using existential quantification over worlds, and so is sensitive only to the position in the world-ordering of the *highest-ranked* worlds in a proposition. For instance, if there is a single ϕ-world u such that $u \succcurlyeq_{\mathbf{g}(w)} v$ for all ψ-worlds v, then $\phi \succcurlyeq^o \psi$ — even if all ψ-worlds strictly dominate all ϕ-worlds except for u.[4]

The lottery example suggests that the orderings that underlie epistemic comparison in English make use of more information than the Lewis/Kratzer semantics allows. Suppose that we know

4 Note that the fact that $\succcurlyeq_{\mathbf{g}(w)}$ and $\succcurlyeq^o$ are not connected does not affect the validity of (25) because the truth of the premises guarantees that we are dealing with connected sub-branches of $\succcurlyeq^o$. However, non-connectedness means that — unlike the modified Lewisian semantics — it is not the case in Kratzer's semantics that the likelihood of a disjunction is always equal to the likelihood of the most likely disjunct, even with finite models or the limit assumption.

that ϕ and ψ are comparable, but we don't know whether ϕ is as likely as ψ or not. According to the Lewis/Kratzer semantics, finding out whether it is only requires us to know the relative positions of the highest-ranked comparable worlds in each. The fact that (25) is not valid for English suggests that we cannot ignore non-maximal worlds: we need information about all the worlds in a proposition. In §2 we will see several ways to spell out this suggestion formally. First, though, we will consider a revised version of Kratzer's theory and see whether it helps with the problem under consideration.

1.4 A modified disjunction puzzle in Kratzer 2012

In a recent book revising and expanding her classic papers on modality and conditionals, Kratzer (2012) suggests that the problem involving disjunction is due to an inessential feature of the theory: "the critical assessments in Yalcin (2010) and Lassiter (2010a) are not yet sufficiently responsive to the important fact that we are very likely to need different notions of comparative possibility to account for different types of comparative modal operators in natural languages" (p.41). Naturally, the proof that the troublesome inference is valid depends on the details of the definition in (20): if we were to lift orderings on worlds to orderings on propositions by a different method, the problem might not arise. It is important for our project, then, to understand how much of the problem under consideration is generated by essential features of the Lewis/Kratzer approach, and how much is a result of the particular lifting rules that they happen to adopt.

Kratzer (2012: 42) suggests a revised definition of comparative possibility, apparently aimed at an empirical objection raised by Yalcin (2010: 922). Yalcin points out that, as a special case of the disjunction puzzle in (11), we can prove that (26a) entails (26c):

(26) a. $\phi \succcurlyeq^o \neg\phi$
b. $\phi \succcurlyeq^o \phi$ (by reflexivity of $\succcurlyeq^o$)
c. $\therefore\ \phi \succcurlyeq^o (\phi \vee \neg\phi)$

Since $(\phi \vee \neg\phi)$ is a tautology, this means that, by the original definition of comparative possibility, $\phi \succcurlyeq^o \neg\phi$ entails $\phi \succcurlyeq^o \top$, as well as $\phi \succcurlyeq^o \psi$ for arbitrary ψ; if it is as likely to rain as it is not to rain, then it is as likely to rain as it is that the sun will rise tomorrow, or that $2+2=4$. The revised definition of comparative possibility does not encounter this difficulty.

(27) Comparative Possibility (revised): ϕ is as good a possibility as ψ (relative to **f**, **g**, and w) if and only if there is no world $u \in (\bigcap \mathbf{f}(w) \cap \psi \cap \neg\phi)$ such that, for every world v in $(\bigcap \mathbf{f}(w) \cap \phi \cap \neg\psi)$, $u >_{\mathbf{g}(w)} v$.

Again, for convenience I will introduce an abbreviation, this time with the superscript ^{n} (for "new").

(28) $\phi \succcurlyeq^n \psi \text{ iff}_{df} \neg\exists u \in (\psi \wedge \neg\phi)\ \forall v \in (\phi \wedge \neg\psi)[u >_{\mathbf{g}(w)} v] \quad (u, v \in \bigcap \mathbf{f}(w))$

This is, of course, equivalent to (29) —

(29) $\phi \succcurlyeq^n \psi \text{ iff } \forall u \in (\psi \wedge \neg\phi) \exists v \in (\phi \wedge \neg\psi)\ [\neg(u >_{\mathbf{g}(w)} v)]$

— again assuming that u and v are in $\bigcap \mathbf{f}(w)$.

The new definition of comparative possibility ranks two propositions ϕ and ψ by ignoring worlds that are in both, comparing only worlds that are in one but not the other. This definition places much weaker requirements on the world-ordering than the original order $\phi \succcurlyeq^o \psi$ did. Previously we required that all of the ψ-worlds be comparable to and weakly dominated by at least one ϕ-world; the new order $\succcurlyeq^n$ requires only that all of the relevant ψ worlds be either (a) weakly dominated by or (b) incomparable with at least one relevant ϕ-world.

The inference in (26), modified to replace $\succcurlyeq^o$ by $\succcurlyeq^n$ throughout, does not go through.

(30) a. $\phi \succcurlyeq^n \neg\phi$

b. $\phi \succcurlyeq^n \phi$

c. $\therefore\ \phi \succcurlyeq^n (\phi \vee \neg\phi)$ **(not valid)**

Essentially, the reason is that the truth-conditions for $\phi \succcurlyeq^n (\phi \vee \neg\phi)$ only take into account the worlds that are in ϕ but not in $\phi \vee \neg\phi$. Since there are none, the only way that $\phi \succcurlyeq^n (\phi \vee \neg\phi)$ can be true is if there are no $\neg\phi$-worlds in $\bigcap \mathbf{f}(w)$. This prediction is quite reasonable: the proof that ϕ is as likely as a tautology goes through only if our background knowledge excludes the possibility of $\neg\phi$.

Before we conclude that the threat has been averted, though, we should consider the form of possible counter-models to (30). The only way that the conclusion can be false when the premises are true is if one of the conditions in (31) holds. It follows immediately from this that a counter-model to the disjunction puzzle must contain worlds either in $(\phi \wedge \chi)$ or in $(\phi \wedge \psi)$, i.e. these three propositions cannot all be disjoint (see Appendix, §6.2 for a proof).

(31) a. $\exists u \in (\psi \wedge \neg\phi)\ \exists v \in (\phi \wedge \chi)[\neg(u \succ_{\mathbf{g}(w)} v)]$

b. $\exists u' \in (\chi \wedge \neg\phi)\ \exists v' \in (\phi \wedge \psi)[\neg(u' \succ_{\mathbf{g}(w)} v')]$

This explains why Yalcin's objection from (26) is avoided by adopting the new definition of comparative possibility: substituting ϕ for χ in (30) guarantees that (unless ϕ is impossible) there will be accessible worlds in $(\phi \wedge \chi) = \phi$. We can then use (31a) to construct a counter-model. The problematic inference pattern in (30) will only be valid if there are more than two disjoint alternatives under consideration.

If the alternatives are disjoint, though, there are no worlds in $(\phi \wedge \chi)$ or in $(\phi \wedge \psi)$, and so no counter-model can be constructed and the inference goes through. This was the case in the lottery scenario that we considered in §1.1: since there can be only one winner in the lottery at hand, if x wins then y and z lose for all $x \neq y \neq z$. So the set of epistemic possibilities in which x_i wins is disjoint from the set in which x_j and x_k win, etc. This means that neither of the conditions in (31) can be satisfied, and it is not possible to construct a counter-model.

A more direct way to make the same point is the following: if we modify the disjunction puzzle by explicitly encoding the fact that we are comparing more than two mutually exclusive propositions, it is valid, with a proof very similar to the proof that the inference was valid in the original version of the disjunction puzzle. For the primary empirical puzzle that we discussed in §1.1, the alternatives are disjoint anyway, and it does not affect the argument if we mention this fact explicitly.

(32) **Modified disjunction puzzle**:

a. $(\phi \wedge \neg\psi \wedge \neg\chi) \succcurlyeq^n (\neg\phi \wedge \psi \wedge \neg\chi)$

b. $(\phi \wedge \neg\psi \wedge \neg\chi) \succcurlyeq^n (\neg\phi \wedge \neg\psi \wedge \chi)$

c. $\therefore (\phi \wedge \neg\psi \wedge \neg\chi) \succcurlyeq^n ((\neg\phi \wedge \psi \wedge \neg\chi) \vee (\neg\phi \wedge \neg\psi \wedge \chi))$

For example,

(33) a. It's as likely that Sam will win (and Mary and Sue won't) as it is that Mary will win (and Sam and Sue won't).

b. It's as likely that Sam will win (and Mary and Sue won't) as it is that Sue will win (and Sam and Mary won't).

c. ∴ It's as likely that Sam will win (and Mary and Sue won't) as it is that either Mary or Sue will win (and Sam won't).

(33) is just as clearly intuitively invalid as the simpler version of the puzzle given in (30): the parentheticals don't add anything, since we already knew that only one individual could win. But we can prove from the definition of $\succcurlyeq^n$ that (32) is a valid argument (see the Appendix, §6.3).

Modifying the ordering on propositions from $\succcurlyeq^o$ to $\succcurlyeq^n$ does not resolve the problem. Both the old and the new versions of comparative possibility yield incorrect predictions about the interaction between epistemic modals and disjunction, and in particular validate the inference from *Sam is as likely to win as anyone else is* to *Sam is as likely to win as he is not to win* in the lottery scenario.

1.5 Theoretical Upshot

These results indicate that Kratzer's (1991; 2012) proposals involving epistemic comparison are incorrect. However, it does not follow that no revision of Comparative Possibility is able to avoid making problematic predictions about disjunction: there are several ways out.

Most obviously, there are many ways to use an ordering on worlds to determine an ordering on propositions. Perhaps if we search this space carefully, we will find an alternative version of Comparative Possibility which allows us to maintain the rest of Kratzer's theory. There is, however, reason to doubt that this approach will be fruitful.

Let $\succcurlyeq_{\mathbf{g}(w)}$ be an ordering on worlds and let $\succcurlyeq^P$ be the order on propositions which we are trying to determine (in conformity with the intuitions of English speakers about entailments among epistemic sentences) using $\succcurlyeq_{\mathbf{g}(w)}$ and some as-yet unknown modification of Comparative Possibility. Suppose that $w_1 \succcurlyeq_{\mathbf{g}(w)} w_2$ and $w_1 \succcurlyeq_{\mathbf{g}(w)} w_3$. Presumably we want to ensure that singletons agree with $\succcurlyeq_{\mathbf{g}(w)}$, and so the revised version of Comparative Possibility should ensure that $\{w_1\} \succcurlyeq^P \{w_2\}$ and $\{w_1\} \succcurlyeq^P \{w_3\}$. We do not, of course, want to require in this situation that $\{w_1\} \succcurlyeq^P \{w_2, w_3\}$, since doing so would lead back to the disjunction puzzle. However, we don't want to rule out the possibility that $\{w_1\} \succcurlyeq^P \{w_2, w_3\}$ *does* hold in some models, either: perhaps $\{w_1\}$ is *much* more likely than either of the others, enough that it is also more likely than their union. But if the order on worlds is to determine a unique order on propositions, we have to make a decision one way or the other — without paying attention to what is happening in the worlds! If we want a determinate answer, while also maintaining the flexibility to go different ways in different cases, we need more information than is contained in a binary order.

The only way out, it seems, is to rule these propositions incomparable in the lifted ordering. Holliday & Icard (2013a) describe a modification of Comparative Possibility which has this feature:

$\succcurlyeq^m$ is a relation which holds between ϕ and ψ whenever it is possible to associate each world $v \in \psi$ with a **unique** $u \in \phi$ such that $u \succcurlyeq_{\mathbf{g}(w)} v$. Tailored to Kratzer's framework the lifting rule looks like this:

(34) $\phi \succcurlyeq^m \psi$ (relative to $\mathbf{f}$, $\mathbf{g}$, and w) iff there is an injection h with domain $\psi \cap \bigcap f(w)$ and range $\phi \cap \bigcap f(w)$, such that, for all $w' \in \psi \cap \bigcap \mathbf{f}(w)$, $h(w') \succcurlyeq w'$.

As Holliday & Icard (2013a) show, a semantics which uses $\succcurlyeq^m$ to interpret *at least as likely as* does not validate (1) and behaves well with respect to a variety of intuitive validities and invalidities (which are also shared by the probabilistic and qualitatively additive semantics discussed below.)

This lifting enforces a good deal of incomparability in the *at least as likely as* relation, because "there is simply not enough information in every total ordering on worlds to completely determine a total likelihood ordering on propositions" (Holliday & Icard 2013a: 522). The situation is even more extreme when the ordering on worlds is not total, of course. I don't know whether this is an empirical defect: epistemic incomparability is thought by many to be a real phenomenon (see Keynes 1921; Fine 1973; Holliday & Icard 2013a for discussion), and the important question is whether the approach makes reasonable predictions about incomparability in the context of concrete scenarios and a specific theory about how the world-ordering relates to the facts of the scenario. In any case, as semantics built around $\succcurlyeq^m$ does quite well logically: the only direct objection I can think of is that it does not explain the acceptability of ratio modifiers with *likely* (see §2.2 below).

While the possibility of modifying Comparative Possibility in this way might seem to resolve the issues with *likely* that we have been discussing, it is something of a Pyrrhic victory for Kratzer's overall framework. Suppose that we modify this theory by replacing $\succcurlyeq^o$ and $\succcurlyeq^n$ by $\succcurlyeq^m$, without changing anything else. Then the following intuitively obvious principle — which is similar to, but even weaker than, the inference in (3) above — comes out invalid.

(35) If ϕ must be the case, then ϕ is more likely than $\neg\phi$.

Here is a minimal counter-model: $W = \{w_1, w_2, w_3\}$, $\mathbf{f}(w_1) = \{W\}$, and $\mathbf{g}(w_1) = \{\{w_1\}\}$. Then *must*$(\{w_1\})$ is true, and we also have $w_1 \succ_{\mathbf{g}(w)} w_2$, $w_1 \succ_{\mathbf{g}(w)} w_3$, and $w_2 \approx_{\mathbf{g}(w)} w_3$. In the modified account we are considering, neither $\{w_1\} \succcurlyeq^m \{w_2, w_3\}$ nor $\{w_2, w_3\} \succcurlyeq^m \{w_1\}$: they are incomparable, and so (35) is not valid. In fact any model in which the set of "best" worlds has cardinality less than $1/2 \times |\bigcap f(w)|$ will generate a counter-example, because there is no injection from $\mathbf{BEST}(\mathbf{f}, \mathbf{g}, w)$ to $\bigcap \mathbf{f}(w) - \mathbf{BEST}(\mathbf{f}, \mathbf{g}, w)$. (This problem can be avoided by adopting a non-Kratzerian semantics for *must*, as Holliday & Icard (2013a) do, and as we will discuss in §4 below.)

A second interpretation of $\succcurlyeq^m$, hinted at by its inventors, does not force us to assume a great deal of epistemic incomparability: we can use $\succcurlyeq^m$ as a method of checking whether a given ordering on worlds is compatible with an antecedently given ordering on propositions. That is, we might have a semantics for *at least as likely as* which determines for most or all pairs of propositions whether one is as likely as the other, and use $\succcurlyeq^m$ to verify that this order is compatible with the ordering $\succcurlyeq_{\mathbf{g}(w)}$ which is used to provide truth-conditions for (e.g.) the epistemic auxiliaries.

(36) If $\phi \succcurlyeq^m \psi$, then ϕ is at least as likely as ψ.

On this account, the incomparability of $\{w_1\}$ and $\{w_2, w_3\}$ in $\succcurlyeq^m$ is interpreted as meaning that the world-ordering does not **constrain** the relative likelihood of these propositions. We may, however, assign a truth-value to this comparison nonetheless.

Below (§3) we will consider this and a number of alternative principles which attempt to constrain the relationship between likelihood and a world-ordering so that logically valid principles can be encoded. In order to do this, though, we need an alternative account of likelihood which does not have the problems that were identified for Kratzer's account. The following section considers several alternatives framed in scalar terms, and the next considers the consequences for Kratzer's semantics for the epistemic auxiliaries.

2 Epistemic comparison without the disjunction puzzle

Let us suspend for the moment the assumption that world-orderings are more basic than proposition-orderings. That is, instead of asking what sort of propositional ordering we can derive given that we have only an ordering on worlds to work with, we will ask what structure an ordering on propositions must have in order to conform with our intuitions about the entailments of epistemic sentences. Later, we will return to the question of whether and how these intuitions can be made to cohere with a semantics for the epistemic auxiliaries built around orderings on worlds.

2.1 Scales of probability

The first solution that we will consider is the one suggested by Yalcin (2007, 2010); Swanson (2006); Lassiter (2010a): a degree theory built on probability. Note that an equivalent formulation of probability in purely qualitative (order-theoretic) terms is possible, as discussed by Savage (1954); Krantz, Luce, Suppes & Tversky (1971); Fishburn (1986); Narens (2007); Lassiter (2011, 2014a); Holliday & Icard (2013a) and many others. The choice between the qualitative and quantitative versions of the theory is analogous to the choice between degree- and order-theoretic versions of the theory of gradation (cf. Cresswell 1976; Klein 1991; Rullmann 1995; van Rooij 2011; Lassiter 2011, 2014a), and does not seem to be important except perhaps in making detailed decisions about how to set up the compositional semantics. I focus on the quantitative version here simply because it is simpler to state and easily integrated with familiar degree-based semantics for scalar adjectives. (Note that other ways of ranking propositions discussed in this paper, including the Lewis/Kratzer account, can be formulated in either qualitative or quantitative terms as well. On a formal level, there is really nothing to choose here, though perhaps there are some interesting philosophical or psychological issues involved. I use the term "scalar" to describe both kinds of theories.)

Probability spaces are standardly defined using standard concepts from possible-worlds semantics, supplemented by measure functions, a notion familiar from degree semantics. We assume a measure function μ which takes propositions (sets of worlds) to real numbers between 0 and 1.[5] We require that a tautology receive measure 1, and that the measure function is *additive* — when two propositions are disjoint the measure of their disjunction is the sum of their individual measures.[6]

5 The choice of 0 and 1 as the minimum and maximum points is arbitrary. With some slight complications in the axiomatization, any two values would do as long as the maximum is strictly greater than the minumum.

6 Note that (37e) requires additivity for finite sets of disjoint propositions only, while the standard possible-worlds definition of probability due to Kolmogorov (1933) is additive for countable disjoint sets. The difference won't matter for our purposes, though it is important in many cases, e.g. when reasoning about real-valued variables.

(37) A **probability space** is a triple $\langle W, \Phi, \mu \rangle$, where

a. W is a set of possible worlds;

b. $\Phi \subseteq \mathcal{P}(W)$ is an algebra of propositions (sets of worlds) containing W which is closed under union and complement;

c. $\mu : \Phi \to [0,1]$ is a function from elements of Φ to real numbers between 0 and 1;

d. $\mu(W) = 1$;

e. **Additivity**: If A and B are in Φ and $A \cap B = \varnothing$, then $\mu(A \cup B) = \mu(A) + \mu(B)$.

Given that *likely* is a gradable adjective, it is natural to frame an account of its semantics in terms of mainstream scalar theories of the semantics of gradable adjectives (cf. Yalcin 2007, 2010; Portner 2009; Lassiter 2010a; Klecha 2012). In Kennedy's (2007) theory of gradability, an adjective such as *heavy* denotes a measure function which maps individuals to their weight, where the latter is construed as a point on a scale which extends from 0 to ∞, exclusive.

(38) $\mu_{weight} : \mathbf{D}_e \to (0, \infty)$

(39) $[\![heavy]\!] = \lambda x_e . \mu_{weight}(x)$

An axiomatization of μ_{weight} would have to encode a condition corresponding to additivity (37e) in order to capture the fact that the weight of an object formed of two smaller objects is the sum of their individual weights. See Krantz et al. 1971; Lassiter 2011, 2014a for discussion.

In Kennedy's theory comparative and equative sentences have the following truth-conditions (by a compositional process whose details are not crucial here):

(40) $[\![x \textit{ is heavier than } y]\!] = 1$ iff $\mu_{weight}(x) > \mu_{weight}(y)$

(41) $[\![x \textit{ is as heavy as } y]\!] = 1$ iff $\mu_{weight}(x) \geq \mu_{weight}(y)$

Similarly, μ_{prob} is a measure function which maps propositions to $[0,1]$ and obeys the constraints in (37). The truth-conditions of comparative likelihood sentences are parallel to those of comparisons of weight, and Kennedy's (2007) compositional apparatus can be ported over in a straightforward way.[7]

(42) $\mu_{prob} : \mathbf{D}_{\langle s,t \rangle} \to [0,1]$

(43) $[\![likely]\!] = \lambda p_{\langle s,t \rangle} . \mu_{prob}(p)$

(44) $[\![\phi \textit{ is more likely than } \psi]\!] = 1$ iff $\mu_{prob}(\phi) > \mu_{prob}(\psi)$

(45) $[\![\phi \textit{ is as likely as } \psi]\!] = 1$ iff $\mu_{prob}(\phi) \geq \mu_{prob}(\psi)$

The problematic disjunction puzzle does not arise in this semantics for *likely*: (46) is not a valid inference.

7 Specifically, the definitions of degree operators such as *pos* and *more* must be modified so as to take arguments of any type $\langle\langle \alpha, d \rangle\rangle$ for α in (at least) $\{e, \langle s,t \rangle\}$. Similar small modifications are sufficient for other degree-based and qualitative theories of comparison as well.

(46) a. $\mu_{prob}(\phi) \geq \mu_{prob}(\psi)$
b. $\mu_{prob}(\phi) \geq \mu_{prob}(\chi)$
c. **not valid**: $\therefore\ \mu_{prob}(\phi) \geq \mu_{prob}(\psi \vee \chi)$

The reason is that the probability of a disjunction of disjoint propositions is (by additivity) the sum of the probabilities of the disjuncts. Let ϕ, ψ, and χ be disjoint propositions, where $\mu_{prob}(\phi) > \mu_{prob}(\psi)$ and $\mu_{prob}(\phi) > \mu_{prob}(\chi)$, and suppose that one of them must be true. Then, as long as $\mu_{prob}(\phi) < .5$ — or, if you like, as long as ϕ is less likely than $\neg\phi$ — then $\mu_{prob}(\psi \vee \chi)$ will be greater than $\mu_{prob}(\phi)$. For example, if we have

- $\mu_{prob}(\phi) = .4$
- $\mu_{prob}(\psi) = .3$
- $\mu_{prob}(\chi) = .3$

then the additivity of probability for disjoint propositions entails

- $\mu_{prob}(\psi \vee \chi) = .6$

and we have a counter-model for the disjunction puzzle: (46a) and (46b) are true but (46c) is false. However, the premises of (46) are compatible with the conclusion: if ϕ had probability .5 or greater, then the probability of $\phi \vee \psi$ could not exceed .5, and (46a)-(46c) would all be true. The probabilistic semantics has the flexibility to go both ways, as desired.

In cases with more than three alternatives such as the lottery example from §1.1, this semantics gives us intuitively reasonable predictions. In a fair lottery (i.e., one in which each ticket is equally likely to be chosen), the probability that individual x will win is given by the number of tickets m that x holds, divided by the total number n of tickets. With $n = 1,000,000$ tickets and no individual holding more than two of them, x's chance of winning is at most 2 in 1,000,000, and x's chance of not winning is vastly greater — to be precise, it is at least 999,998 in 1,000,000, the sum of the chances of winning of all individuals other than x. This accords very closely with our intuitions about likelihoods of events in the lottery scenario.[8]

A probabilistic scalar semantics for *likely* delivers intuitively correct predictions about the disjunction puzzle, and does so in a way which does not require adding additional theoretical mechanisms in formal semantics beyond what is required for the semantics of scalar expressions in general. This constitutes a strong argument favoring the probabilistic scalar semantics for *likely* over an approach built directly upon Comparative Possibility. However, it does not show decisively that probability provides the right scale.

8 Portner (2009) argues that this semantics is problematic because the probability scale is upper-bounded, and so — on Kennedy & McNally's (2005) assumptions — *ϕ is completely likely* ought to entail that $\mu_{prob}(\phi) = 1$. Various responses are available: for instance, Lassiter (2010a) suggests that this data point constitutes an argument against Kennedy & McNally's (2005) interpretation of *completely*, and adduces other counter-examples. Klecha (2012) argues that Portner's observation shows that the probability measure that *likely* denotes is lexically restricted to exclude the endpoints from its domain, along the lines of Kennedy's (2007) proposal for *expensive*. On Kennedy & McNally's theory, we would expect this change to prevent *completely likely* from having a degree-maximizing reading.

2.2 Alternative scalar approaches

Given that a semantics build around probability measures fails to validate the offending inference, a weaker degree semantics which allows — but does not require — probability measures will of course fail to validate the inference as well. For example, a semantics built around symmetric fuzzy measures has this property (and — as Holliday & Icard (2013b) show — a number of other desirable properties identified by Yalcin (2010)). Let W be a set of worlds and $\Phi \subseteq \mathcal{P}(W)$ be a set of propositions closed under complement:

(47) A **symmetric fuzzy measure** is a function $\mu : \Phi \to [0,1]$ such that

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

   b. **Symmetry**: $\mu(A) + \mu(-A) = 1$ for all $A \in \Phi$;

   c. **Monotonicity**: If $A \subseteq B$, then $\mu(A) \leq \mu(B)$ for all $A, B \in \Phi$.

Every probability measure is a symmetric fuzzy measure, and so the counter-model given in the last section suffices to show that (1) is not valid for symmetric fuzzy measures either. However, many symmetric fuzzy measures are not probability measures: the crucial difference is that probability measures are additive, while symmetric fuzzy measures obey the weaker condition of monotonicity.

A semantics built around symmetric fuzzy measures would resolve the disjunction puzzle, but also fail to validate intuitively valid inferences. Suppose that Sam might go to school, but he is somewhat more likely to cut class and go to the movies. If *likely* denotes a symmetric fuzzy measure, then this scenario is compatible with (48):

(48) It is exactly as likely that Sam will go to the movies as it is that he will either go to school or go to the movies.

This is clearly wrong. Unless it is impossible that Sam will go to school, it is more likely that he will go to school *or* the movies than it is that he will go to the movies. We can fix this problem by replacing the monotonicity axiom (47c) with a stronger condition.

(49) **Equal shares**: $\mu(A) \geq \mu(B)$ if and only if $\mu(A - B) \geq \mu(B - A)$.

This axiom requires that the portions of logical space in which A and B are both true contribute an equal measure to each. This condition rules out the odd situation just described: $\mu(\textbf{movies}) \geq \mu(\textbf{movies} \cup \textbf{school})$ iff $\mu(\textbf{movies} - [\textbf{movies} \cup \textbf{school}]) \geq \mu([\textbf{movies} \cup \textbf{school}] - \textbf{movies})$. Since $\textbf{movies} - [\textbf{movies} \cup \textbf{school}] = \varnothing$, this is equivalent to the condition that $\mu(\varnothing) = 0$ is greater than or equal to $\mu(\textbf{school})$. (48) thus entails that $\mu(\textbf{school}) = 0$ on the updated semantics, as desired.

Adding the condition in (49) to (47) gives us the semantics based on **qualitatively additive** measures that is discussed by Holliday & Icard (2013a: §7). This semantics is also weaker than the probabilistic semantics discussed in §2.1, but it maintains a number of desirable logical properties — including, of course, the invalidity of (1). Holliday & Icard argue persuasively that it will be difficult, perhaps impossible, to distinguish this semantics from the probabilistic one described above on the basis of intuitive entailments among the items in their language (those that are rendered in English as *likely*, *as likely as*, *more likely than*, *epistemically possible*, and *epistemically necessary*).

This conclusion leaves open the possibility that an argument could be constructed in favor of the probabilistic semantics, using a larger fragment of English and theoretical insights from other areas of natural language semantics. Here is one argument along these lines. Imagine that you are about to throw a fair die. The following sentence is intuitively true and assertible:[9]

(50) You are exactly three times as likely to throw an odd number as you are to throw a two.

Or, again, (51) is reasonable if you are about to draw a card at random from a well-shuffled pack:

(51) You are exactly twice as likely to draw a jack as you are to draw a red queen.

(52) gives some similar examples found on the web.

(52) a. In 1950, black families were over twice as likely to be in the lowest fifth of income than white families; by 1977, they were exactly twice as likely as whites to be in the lowest family income quintile.
b. [A]ny given blue output is exactly 4 times as likely as any given green output. (I.e., 20% of the time on average you will receive a green back and 80% of the time a blue.)

In the literature on degree modification, it has been argued that ratio modifiers are acceptable with an adjective only if its scale is additive (or satisfies equivalent qualitative conditions: see Sassoon 2010; van Rooij 2011; Lassiter 2011, 2014a). The intuitive contrast between (53) and (54) illustrates:

(53) ✓ Mary is exactly three times as tall/old/heavy as Bill is.

(54) ?? Mary is exactly three times as angry/hungry/lecherous as Bill is.

It is easy to imagine situations in which (53) could be used, but rather difficult for (54). With rare exceptions, examples of this form found on the web involve adjectives that fall onto scales with additive measures, such as *long, old, much, many, wide, big, fast, frequent, often, massive*, and of course *likely*.[10]

---

9 Thanks to Alan Bale for emphasizing the importance of using precise measurements, as enforced by *exactly*, in this argument.

10 Less obvious examples found in the first 100 Google hits include *risky*, *bright* in an astronomical context, and *invasible* in reference to the probability that a species will succeed in invading a new habitat. I suspect that additivity can be motivated for these adjectives' scales as well.

Many apparent counter-examples turn out to be cases in which an adjective is re-interpreted as belonging to a scale which is additive. Here are some examples from the web. (1) indicates that power is to be measured in megatons. (2) humorously suggests that funniness can be measured in the number of exploding golf balls purchased.

(1) The largest bomb now in the active arsenal of the US is said to be 1.2 megatons, which is almost exactly 100 times as powerful as the Hiroshima bomb.

(2) The Exploding Golf Ball Four Pack is exactly four times as funny as the regular Exploding Golf Ball, and that's wicked funny!

There are also examples such as (3), where the author plays on the fact that the measure phrase *halfway* occurs in the idiom *halfway decent*; the humor is generated precisely by the fact that degrees of decency are not the right sort of object to be measured using ratio or proportional modifiers.

(3) Rosen is the U.S. Attorney for D.C., and he is, in fact, a halfway decent person. This makes him exactly four times as decent as Olivia and eight times as decent as Fitz. (And 10 or 12 times as decent as Mellie, but far less awesome.)

(55) a. [T]he opening note lasted exactly three times as long as the second one ...

b. A strain of pathogenic bacteria undergoes cell division exactly two times as often as a nonpathogenic mutant strain of the same bacteria.

If this constraint on the acceptability of ratio modifiers is correct, we can infer that degrees of likelihood are additive, as in the probabilistic semantics. This conclusion is not in conflict with Holliday & Icard's (2013a) arguments, for two reasons. First, they explicitly restrict attention to a language without modified forms of *likely* other than comparatives and equatives. Second, their argument is about what can be inferred exclusively from data involving valid and invalid inferences. The argument just mentioned relies instead on theoretical claims about the structure of the models in which expressions are interpreted, and in particular the empirical signature of adjectival scales with certain formal properties.

In what follows I will consider the implications of adopting a semantics build around finitely additive probability for the semantics of the epistemic auxiliaries. Naturally, the strength of the argument for privileging finite additivity over qualitative additivity is tied to the strength of the argument that ratio modifiers are acceptable with adjectives in English just in case if the adjective's scale is (fully) additive. If this general claim about degree modification and scalar structure turns out to be incorrect, it may be advisable to reconsider the weaker semantics built around qualitatively additive measures. It would then be even more important to answer a question which I do not address further here: what are the implications of adopting a (merely) qualitatively additive semantics for a Kratzerian account of the epistemic auxiliaries?

3 Implications for epistemic auxiliaries

Let us suppose, then, that *likely* is a scalar adjective whose scale satisfies the axioms of finitely additive probability listed in (37). A conservative response to this conclusion might be to grant that *likely* and *probable* have a complex logic of their own, perhaps one built around probability, and abandon the ambition of using an order on worlds to derive an appropriate order on propositions. This move might make it possible to maintain Kratzer's attractive semantics for *must* and *might* while resolving the disjunction puzzle.

The need to account for logical relations between likelihood and other epistemic modals places constraints on this effort: there must be some kind of formal connection between the ordering over

---

Examples of *exactly n times as Adj as* with adjectives whose default scale is not additive seem to be of one of these two types. This seems to confirm the generalization that additivity is required for ratio modifiers to be used with a precise, non-metaphorical meaning.

Examples without *exactly* occur with somewhat greater frequency with adjectives on clearly non-additive scales, and are clearly not unintelligible. In general, a hyperbolic interpretation seems to be intended (*twice as angry as* = "a whole lot angrier"), or a humorous effect is intended.

(4) Now she felt two times as hungry as she was when she sat up.

(5) A: I was as angry and outraged over the 9/11 attacks as any American.
B: Bull. I was ten times as angry and four times as outraged as you were. And that's nothing. My friend Ernie was three times as angry as I was. (Interestingly, he was only 75% as outraged.)

propositions given by a probability space and the various orderings over worlds and propositions that Kratzer's theory trades in. In this section I discuss several possible ways to combine the two and demonstrate that they either miss clearly valid entailments or validate inference patterns that are not actually valid. This does not show that a hybrid theory is impossible, since we will not be able to consider all possible ways of drawing this connection; but I hope that the discussion will make clear the formal challenge that stands in the way of maintaining even a considerably modified version of the standard theory in light of the disjunction puzzle.

The simplest imaginable effort to connect the Kratzerian semantics with probability would be to assume no particular connection: we have probability spaces, modal bases, and ordering sources coexisting peacefully in a model, free to combine in any way. Probability spaces provide truth-conditions for sentences with *likely* and *probable*, and modal bases and ordering sources help determine truth-conditions for sentences with epistemic *might* and *must*.

This won't work, though, because of the existence of logical connections between the epistemic adjectives and auxiliaries. For example, (56) is clearly a valid argument (cf. §1.5).

(56) a. Sam must be at home.
b. $\therefore$ It is much more likely that Sam is at home than it is that he is not at home.

There is nothing to enforce this entailment in a theory of the type just envisioned, or other similar relations between the epistemic adjectives and auxiliaries. What we need are *bridging rules* which relate the probabilistic ordering over propositions that undergirds *likely* with the ordering over propositions on which — according to the newly restricted Kratzerian theory — *must* and *might* rely. Otherwise, we could construct monstrous models in which, for example, Sam must be at home but it is extremely unlikely that he is.[11]

The first bridge to consider between Kratzer's theory of the auxiliaries and a probabilistic semantics for *likely* is suggested in a somewhat different context by Kratzer (2012).

(57) **BR1**: If $\phi \succcurlyeq^n \psi$, then $\mu_{prob}(\phi) \geq \mu_{prob}(\psi)$.

Note that Kratzer doesn't propose **BR1** as a semantic rule, but as a way to relate what she considers to be the semantically relevant concept of comparative possibility to the semantically irrelevant but scientifically interesting concept of probability. Given the preceding discussion, though, I will focus on the possibility of a semantic use.[12]

11 A referee suggests that Sherlock Holmes' well-known adage is a counter-example to this inference: "How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?" (Arthur Conan Doyle, *The Sign of the Four*, ch.6). I think not. Sherlock's point is about the dynamics of information: after gathering new information (at time t) which eliminates some things that were possible (at time $t' < t$, before the new information was gathered), whatever has not been eliminated must (at t) be true, however improbable it was (at t'). The quote strikes me as a rather nice intuitive gloss on Bayesian inference, and eminently compatible with the validity of (56) as far as an information state at a particular point in time is concerned.

12 Kratzer (2012) denies that probability plays any role in the semantics: "[o]ur semantic knowledge alone does not give us the precise quantitative notions of probability and desirability that mathematicians and scientists work with" (p.25). She proposes a slightly more restrictive version of BR1 as a way to use the comparative possibility concept to provide "conceptual launch pads for mathematical explorations to take off from" (ibid.).

In addition to the problems pointed out in the main text for the semantic application, probability measures which agree with $\succcurlyeq^n$ in this way will have quite limited applicability for the non-semantic purpose that Kratzer intends. The

Kratzer (2012: 42-3) suggests this rule as a bridge between probability and Comparative Possibility, and gives a small model spelling out an example of a $\succcurlyeq^n$ and a μ_{prob} that satisfy this constraint. While there is no technical problem with this formulation, it cannot be the right bridging rule because it would re-introduce the (modified) disjunction puzzle, destroying the advantages of moving to probability in the process. That is, if we adopt (57) as a constraint on models in a semantics for English, the following inference will be valid whenever ϕ, ψ, and χ are disjoint propositions which are ordered by $\succcurlyeq^n$:[13]

(58) a. $\mu_{prob}(\phi) \geq \mu_{prob}(\psi)$
b. $\mu_{prob}(\phi) \geq \mu_{prob}(\chi)$
c. $\therefore$ $\mu_{prob}(\phi) \geq \mu_{prob}(\psi \vee \chi)$

BR1 limits us to considering probability measures which assign ϕ a probability at least as great the probability assigned to $(\psi \vee \chi)$ when ϕ as likely as each of these individually. Adopting a bridging rule which restricts attention to probability measures for which the disjunction puzzle is a valid inference would not be advisable, for reasons discussed in detail above (§1.1). (Note that the toy model that Kratzer (2012: 46) describes happens to have a structure for which this inference is harmless, but this is a special feature of this example which does not generalize to most choices of **g** and μ_{prob}.)

In addition to validating intuitively invalid patterns of inference, **BR1** is too weak in that it fails to validate (56). To see the problem, recall that the truth-conditions of *must*(ϕ) and *might*(ϕ) are given by universal quantification over top-ranked worlds, when these exist. Suppose that $\mathbf{BEST}(\mathbf{f},\mathbf{g},w)$ (as defined in (18)) exists. Clearly this set strictly dominates its own complement in the ordering $\succcurlyeq^n$. This means, by **BR1**, that it must have strictly greater probability. But this is the only constraint that **BR1** imposes: there is no requirement that it must have *much* greater probability, and **BR1** would allow, for example, models in which $\mathbf{BEST}(\mathbf{f},\mathbf{g},w)$ has probability 0.5001 and $W - \mathbf{BEST}(\mathbf{f},\mathbf{g},w)$ has probability .4999. But since *must*$(\phi) = 1$ iff ϕ is true in all worlds in $\mathbf{BEST}(\mathbf{f},\mathbf{g},w)$, this means that *must* ϕ can be true even when ϕ is just barely more likely than its negation. As a result, (56) is not a valid inference as far as the hybrid proposal under consideration is concerned.

---

probabilistic models used in scientific practice will fail to agree with virtually any non-trivial $\succcurlyeq^n$. The reason is that, regardless of your philosophy of probability, probabilities are meant to track the relative frequencies of events in one way or another. (If you're a frequentist, this is a truism; if you're a Bayesian, it holds once you've observed enough data; etc. See e.g. Hacking 2001; Mellor 2005 for discussion of this point.) Relative frequencies are not, however, distributed in the way that BR1 would require. In particular, the counterpart of the (modified) disjunction puzzle clearly does not hold of relative frequencies: an event that is more frequent than each of a set of disjoint alternatives will often turn out to be less frequent than the disjunction of these alternatives. For example, from the fact that drunk driving accidents are pairwise more frequent in Texas than in each other U.S. state we cannot conclude that drunk driving accidents are more frequent in Texas than in the whole United States excepting Texas. This mismatch does not mean that there is anything wrong with adopting the construal of probability under discussion, on a technical level; it just means that this concept of probability would be of quite limited use, since very few useful probability measures will agree with any $\succcurlyeq^n$.

13 The caveat "which are ordered by $\succcurlyeq^n$" is required because **BR1** does not impose any constraints on the probability assigned to a proposition if it does not appear anywhere in the relation $\succcurlyeq^n$. The only way that ϕ could fail to appear somewhere in $\succcurlyeq^n$, though, is if there are not any ϕ-worlds in $\bigcap \mathbf{f}(w)$. Since this entails that ϕ is epistemically impossible, relying on models with this property would probably not be of much help for someone who wishes to defend **BR1**. I will leave out this caveat in what follows.

These problems arise equally if **BR1** is modified by replacing $\succcurlyeq^n$ with $\succcurlyeq^o$: the disjunction puzzle is reintroduced, and (56) is not valid. These are just two obvious strategies for relating probability to the orderings on worlds that Kratzer's semantics relies on, but they serve to illustrate the problem: *prima facie* reasonable proposals can turn out to have unacceptable consequences upon careful inspection.

Another natural suggestion is to start with the ordering on worlds $\succcurlyeq_{\mathbf{g}(w)}$, and — rather than defining a lifted "Comparative Possibility" ordering on propositions as Lewis and Kratzer do — constrain the probability measure to agree with the world-ordering $\succcurlyeq_{\mathbf{g}(w)}$ on singletons.

(59) **BR2**: If $u \succcurlyeq_{\mathbf{g}(w)} v$, then $\mu_{prob}(\{u\}) \geq \mu_{prob}(\{v\})$.

This method of ensuring consistency between a world-ordering and a probability measure is not implausible (with finite W, at least); but it is of little use if the goal is to save Kratzer's theory of the epistemic auxiliaries. When combined with the definitions of *must* and *might* given by Kratzer (1991), such a semantics would fail to validate not only (56), but even the inference from *must* ϕ to ϕ *is more likely than* $\neg\phi$. Suppose that $W = \{w_1, w_2, w_3\}$, $\mathbf{f}(w_1) = \{W\}$, $\mathbf{g}(w_1) = \{\{w_1\}\}$, $\mu_{prob}(\{w_1\}) = .4$, and $\mu_{prob}(\{w_2\}) = \mu_{prob}(\{w_3\}) = .3$. Then (59) is satisfied because $\succcurlyeq_{\mathbf{g}(w_1)}$ and μ_{prob} agree: $w_1 \succcurlyeq_{\mathbf{g}(w_1)} w_2$, $w_1 \succcurlyeq_{\mathbf{g}(w_1)} w_3$, and $w_2 \approx_{\mathbf{g}(w)} w_3$. $\mathbf{BEST}(\mathbf{f}, \mathbf{g}, w) = \{w_1\}$, and so *must*$(\{w_1\})$ is true at w_1; but $\mu_{prob}(\{w_2, w_3)\}) = .3 + .3 = .6$, which is greater than $\mu_{prob}(\{w_1\})$. In this model, *must* ϕ is true but ϕ *is more likely than* $\neg\phi$ is false. Bridging rule **BR2** is thus not an appropriate way to combine a Kratzerian semantics for *must* with a probabilistic semantics for *likely*.

A third option along these lines is drawn from Holliday & Icard 2013a: §9, which was discussed already in §1.5. As a reminder, the modified lifting rule is:

(60) $\phi \succcurlyeq^m \psi$ (ϕ is at least as good a possibility as ψ, relative to $\mathbf{f}$, $\mathbf{g}$, and w) iff there is an injection h with domain $\psi \cap \bigcap f(w)$ and range $\phi \cap \bigcap f(w)$, such that, for all $w' \in \psi \cap \bigcap \mathbf{f}(w)$, $h(w') \succcurlyeq w'$.

We already saw that it would be unwise to combine a qualitative semantics based on this lifting with Kratzer's proposal for *must*. Will a weaker bridging principle based on this lifting do better?

(61) **BR3**: If $\phi \succcurlyeq^m \psi$, then $\mu_{prob}(\phi) \geq \mu_{prob}(\psi)$.

No: in the counter-model to **BR2** considered just above, *must*$(\{w_1\})$ is true, and we also have $w_1 \succ_{\mathbf{g}(w)} w_2$, $w_1 \succ_{\mathbf{g}(w)} w_3$, and $w_2 \approx_{\mathbf{g}(w)} w_3$. The inference from *must* ϕ to ϕ *is more likely than* $\neg\phi$ fails, for example, in the situation in which $\mu_{prob}(\{w_1\}) = .4$ and $\mu_{prob}(\{w_2\}) = \mu_{prob}(\{w_3\}) = .3$. Since $\{w_1\}$ and $\{w_2, w_3\}$ are incomparable in $\succcurlyeq^m$, this model is compatible with BR3.

Perhaps some alternative method of lifting an ordering on worlds to an ordering on propositions could be devised that both avoids empirical problems and coheres with Kratzer's semantics for the auxiliaries, but the prospects do not look very good at present. Alternatively, it might be possible to use a modal base and ordering source to constrain assignments of probability without using a world-ordering as an intermediary, e.g., by including in the ordering source explicit constraints on probability assignments.

In the remainder of this paper, I will suggest a more straightforward alternative: we can look for a semantics for the epistemic auxiliaries which relates in a well-behaved way to probabilistic concepts, circumventing (or at least trivializing) the modal base/ordering source apparatus. The next section considers some details of the semantics of *must* and *might* if we take this route.

4 Probability and the epistemic auxiliaries

The difficulties encountered in reconciling Kratzer's semantics for the auxiliaries with a probabilistic semantics for *likely* can be resolved by taking two steps. The first is to stipulate that epistemic interpretations of the auxiliaries always make use of an empty ordering source: $\mathbf{g}(w) = \varnothing$ for all w. In this case, all worlds compatible with the modal base propositions are "best" worlds, and Kratzer's *must* and *might* collapse into universal/existential quantifiers over a set of "epistemically accessible" worlds, $\bigcap \mathbf{f}(w)$. This semantics is thus a variant of usual treatment of epistemic modals in modal logic, with the primary difference being that we derive the domain of quantification in a less direct way.

The second step is to ensure that the domain of μ_{prob} is an algebra generated by this same set of worlds: each world w is associated with some μ_{prob} which is a function from $\Phi \subseteq \mathcal{P}(\bigcap \mathbf{f}(w))$ to the $[0,1]$ interval, where Φ is closed under union and complement. If we take both of these steps, we predict that *must* ϕ entails $\mu_{prob}(\phi) = 1$, and that $\mu_{prob}(\phi) > 0$ entails *might* ϕ. Clearly, (56) will come out valid in this case: If *must* ϕ is true, then $\mu_{prob}(\phi)$ is greater than $\mu_{prob}(\neg\phi)$ by the largest possible margin. A semantics along these lines enforces strong and generally plausible logical connections between epistemic comparatives and auxiliaries, as Yalcin (2010) details.[14]

Another way to approach the issue would be to simply define *must* using probabilities: *must* ϕ is true iff $\mu_{prob}(\phi) \geq \theta$ for some possibly context-sensitive threshold θ. Assuming *might* is *must*'s dual, we also predict that *might* ϕ is true iff $\mu_{prob}(\phi) > 1 - \theta$. It is useful to distinguish two versions:

- "Strong": $\theta = 1$ (Yalcin 2005)
- "Weak": θ is high, but not necessarily 1 (Swanson 2006; Lassiter 2011)

If $\theta = 1$, this semantics is equivalent to a quantificational treatment of *must* if the domain is finite and includes all and only worlds with non-zero probability. However, as Yalcin (2007) points out, the quantificational and strong probabilistic semantics make different predictions when continuous variables are under discussion (and so uncountable W). If Sam is about to attempt the long jump, *Sam might jump π meters* could well be true, but the probability of this proposition will generally be zero. The quantificational semantics allows that this statement may be true, but the probabilistic semantics does not.

This argument seems to favor the quantificational account, but it relies on a questionable assumption about the interpretation of numerical expressions. It is generally thought that expressions making reference to degrees and numbers are always interpreted with a certain granularity, which may well never reach zero except in mathematical discussions (cf. Lewis 1979; Krifka 2007; Sauerland & Stateva 2007; Bastiaanse 2011). On this account, *Sam might jump π meters* is interpreted as *Sam might jump $\pi \pm g$ meters* for some granularity $g > 0$. The directly probabilistic semantics for *might* and *must* then remains viable: even if *Sam jumps π meters* has probability zero,

14 All of the definitions for *must* and *might* considered in this paper fail to encode the fact that these modals have some characteristics of indirect/inferential evidentials: *It must be raining* is infelicitous if the speaker has directly observed that it is raining. See Palmer (1979); von Fintel & Gillies (2010) for discussion. I will assume, optimistically, that this component of the meaning can be added to any of the proposals under discussion as a presupposition or conventional implicature.

Sam jumps $\pi \pm g$ *meters* will receive non-zero probability in many probability distributions, for many granularities g. This account is also compatible with the weak semantics in which probability 1 is not needed for *must*, but whether the predictions are reasonable in a particular case will depend on the details of the probability distribution, the granularity, and the setting of θ.

Since the argument from continuous sample spaces is not decisive, it may seem that the quantificational and strong probabilistic semantics are for all intents and purposes equivalent, and that both these and the weak probabilistic account are viable. I'll conclude by mentioning three arguments that may help in choosing between these options, without reaching a definitive conclusion.

First: since Karttunen (1972) it has been widely thought that a speaker who utters "That must be Sam" is less committed to the truth of *That is Sam* than someone who simply utters "That is Sam". This empirical claim is repeated by Kratzer (1991), and encoded in her proposal by the fact that *must* quantifies, not over all epistemically possible worlds, but only over those selected by the ordering source as the "best" ((17)-(19)).

If Karttunen is correct, it would seem that both the quantificational and the strong probabilistic semantics that we are considering are too strong: both enforce the condition that *must p* indicates maximal confidence in the truth of p. However, von Fintel & Gillies (2010) have recently argued that that *must* is in fact strong, suggesting that Karttunen's observation can be attributed to an evidential component rather than any truth-conditional weakness. They adduce a number of empirical and theoretical arguments in favor of the strong semantics. Lassiter (2014b) defends the "weakness" claim, pointing to examples drawn from online genealogy discussions and other sources which appear to counter-exemplify the claim that *must* entails maximal commitment.

(62) [T]he 1880 census shows her living with mom, two brothers, and her daughter ... So David [the father] must have died before 1880.

(63) Goodman was still alive in mid-January 1621..., although not in good physical shape. He is listed as one of those who received land in 1623.... He is not listed among those who were part of the cattle division of 1627, so he must have died by then.

Lassiter argues (following Stone 1994) that the propositional argument of *must* is presented as the best explanation of the evidence available to the author, but not as the only epistemically possible option: a marital split and a change of occupation, respectively, would be fairly obvious alternatives. He suggests that a weak probabilistic semantics for *must* is needed if we wish to explain the frequent use of *must* along these lines without attributing massive error to speakers. I will not attempt to resolve this debate here, but will simply note that its resolution will be crucial in the choice among the three theories of the epistemic auxiliaries under consideration.

Second: the quantificational and strong probabilistic semantics both validate the inference in (64), but the weak probabilistic semantics does not.[15]

(64) a. ϕ is more likely than ψ.
  b. $\therefore$ ϕ might be the case.

15 Thanks to a *Journal of Semantics* reviewer for convincing me that this pattern may not be valid, contrary to earlier inclination, and for providing the example in (65).

This pattern seems unobjectionable at first glance. However, consider (65), uttered in a year in which Yale is a much better team than Harvard:

(65) a. Yale is more likely to win the NCAA basketball championship than Harvard is.
b. ∴ Yale might win the NCAA basketball championship.

It is not obvious that (65b) can reasonably be inferred from (65a), given the background knowledge that the odds against either team winning are simply astronomical.

If (64) is indeed valid, the quantificational and strong probabilistic semantics are to be preferred. If the instance of this inference in (65) is not acceptable, though, then we have an argument in favor of the weak probabilistic semantics. This theory is able to accommodate the possibility that a very improbable event is more probable than another, even while both fall short of the $1 - \theta$ threshold. Further empirical work is needed to clarify this issue.

Third: a number of authors have recently observed a connection between the appropriate use of epistemic auxiliaries and the concept of *awareness* (Franke 2007; Franke & de Jager 2011; de Jager 2009; Yalcin 2011). Borrowing Yalcin's example, "Sam believes that it might be raining in Topeka" is not intuitively true simply because Sam has never considered the question of whether it's raining in Topeka. Rather, the statement seems to presuppose that Sam is actively aware of the issue of whether it's raining in Topeka. Yalcin's solution involves (a) a revised semantics for *believes*, and (b) modeling an agent's (non-)awareness of the rain-in-Topeka possibility using a partition over the epistemically accessible worlds. *Might* is then defined as a quantifier over cells of this partition (cf. Yalcin 2007: §7). This semantics is very attractive in its ability to deal with this problem. However, it is also problematized by (62)-(63), unless these can be explained in some other way. Franke & de Jager's (2011) somewhat more powerful theory may be able to cope with these examples, but the details remain unclear, and I suspect that it would still have difficulty with examples like (65) in the current formulation.

Both the strong and weak probabilistic approaches to *must* and *might* also encounter the problem that Yalcin points out: Sam might be in an epistemic state which is maximally non-committal about rain in Topeka, in which case a probability of .5 would be reasonable. Nevertheless, unless he is actively considering the possibility of rain in Topeka, it seems unreasonable to say that he believes that it might be raining there, even though he assigns this proposition a probability greater than any reasonable threshold for *might*. A major desideratum for future work in this vein will be to integrate the probabilistic semantics for the epistemic auxiliaries with a good treatment of belief and awareness.

5 Conclusion

The disjunction puzzle may turn out to be the Achilles' heel of the currently dominant theory of epistemic modality in linguistics. I have argued that this problem is not easily resolved by tweaking the definition of comparative possibility, as Kratzer (2012) has suggested. Nor is the puzzle merely evidence that the adjectival epistemic modals *likely* and *probable* are different from other epistemic modals; we need a semantics that predicts the various logical connections between these items and the epistemic auxiliaries *might* and *must*, and there is no obvious way to do this while maintaining the basic structure of Kratzer's theory.

I argued that a straightforward way to capture the meanings of *likely* and *probable* while avoiding the disjunction puzzle was to assign them a scalar semantics based on ordinary probability. We can formulate this account without adding any new machinery to the semantics of English beyond what is already needed to capture the semantics of gradable adjectives such as *heavy*, *tall* and *full*. Taking this route costs us nothing in terms of the overall complexity of a semantic theory for English. Other scalar accounts are also possible, including the semantics based on qualitatively additive measures analyzed in some detail by Holliday & Icard (2013a). Degree modification data give us a good (but perhaps not yet decisive) argument in favor of the probabilistic semantics.

Even though the probabilistic scalar semantics can be formulated in a fairly conservative way, it has knock-on effects throughout the theory of epistemic modality: the need to capture logical relations between the epistemic auxiliaries and adjectives places considerable pressure on efforts to maintain the basic structure of the Kratzerian theory of *must* and *might* while adopting a probabilistic semantics for *likely*. I considered several possible accounts along these lines, showing that they fail to capture a certain obvious entailment between *must* and *more likely than*.

The most promising approach in light of this discussion seems to be to jettison the modal base/ordering source entirely in favor of an approach known to generate appropriate entailments. A variant of the $\Box/\Diamond$ treatment familiar from pre-Kratzerian modal logic integrates neatly with the probabilistic semantics, as do two theories — one "strong" and one "weak" — in which *must* and *might* constrain probability assignments directly. Careful empirical and theoretical work will be needed in order to choose between these options, including — but probably not limited to — resolutions of the three open problems discussed in the last section.

References

Bastiaanse, Harald. 2011. The rationality of round interpretation. In U. Sauerland R. Nouwen, R. van Rooij & H.-C. Schmitz (eds.), *Vagueness in communication*, 37–50. Springer.

Cresswell, Max J. 1976. The semantics of degree. In B. Partee (ed.), *Montague grammar*, 261–292. Academic Press.

Fine, Terrence L. 1973. *Theories of Probability: An Examination of Foundations*. Academic Press.

von Fintel, Kai & Anthony Gillies. 2010. Must... stay... strong! *Natural Language Semantics* 18(4). 351–383.

Fishburn, Peter C. 1986. The axioms of subjective probability. *Statistical Science* 1(3). 335–345.

Franke, Michael. 2007. The relevance of awareness. In Paul Dekker Maria Aloni & Floris Roelofsen (eds.), *Proceedings of the sixteenth amsterdam colloquium*, 97–102.

Franke, Michael & Tikitu de Jager. 2011. Now that you mention it. In *Language, games, and evolution*, 60–91. Springer.

Goble, Lou. 1996. Utilitarian deontic logic. *Philosophical Studies* 82(3). 317–357.

Hacking, Ian. 2001. *An introduction to probability and inductive logic*. Cambridge University Press.

Halpern, Joseph Y. 1997. Defining relative likelihood in partially-ordered preferential structures. *Journal of Artificial Intelligence Research* 7. 1–24.

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

Hamblin, Charles L. 1959. The modal "probably". *Mind* 68(270). 234–240.

Holliday, Wesley H. & Thomas F. Icard. 2013a. Measure semantics and qualitative semantics for epistemic modals. In Todd Snider (ed.), *Semantics and Linguistic Theory (SALT) 23*, Ithaca, NY: CLC Publications.

Holliday, Wesley H. & Thomas F. Icard. 2013b. On the logic of comparative likelihood. Manuscript, University of California at Berkeley and Stanford University.

de Jager, Tikitu. 2009. *"now that you mention it, i wonder..."': Awareness, attention, assumption*: dissertation.

Karttunen, Lauri. 1972. Possible and must. In J. Kimball (ed.), *Syntax and semantics*, vol. 1, Seminar Press.

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

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

Keynes, John Maynard. 1921. *A Treatise on Probability*. Macmillan.

Klecha, Peter. 2012. Positive and conditional semantics for gradable modals. *Proceedings of Sinn und Bedeutung 16*.

Klein, Ewan. 1991. Comparatives. In A. von Stechow & D. Wunderlich (eds.), *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung*, Walter de Gruyter.

Kolmogorov, Andrey. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Julius Springer.

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

Kratzer, Angelika. 1981. The notional category of modality. In Eikmeyer & Rieser (eds.), *Words, worlds, and contexts*, 38–74. de Gruyter.

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

Kratzer, Angelika. 2012. *Modality and Conditionals: New and Revised Perspectives*. Oxford University Press.

Krifka, Manfred. 2007. Approximate interpretation of number words: A case for strategic communication. In Gerlof Bouma, Irene Krämer & Joost Zwarts (eds.), *Cognitive foundations of interpretation*, 111–126. Koninklijke Nederlandse Akademie van Wetenschapen.

Lassiter, Daniel. 2010a. Gradable epistemic modals, probability, and scale structure. In Nan Li & David Lutz (eds.), *Semantics & Linguistic Theory (SALT) 20*, 197–215. CLC Publications.

Lassiter, Daniel. 2010b. The Algebraic Structure of Amounts: Evidence from Comparatives. *Interfaces: Explorations in Logic, Language and Computation* 38–56.

Lassiter, Daniel. 2011. *Measurement and Modality: The Scalar Basis of Modal Semantics*: New York University dissertation.

Lassiter, Daniel. 2014a. *Measurement and Modality: The Scalar Basis of Modal Semantics*. Oxford University Press (to appear).

Lassiter, Daniel. 2014b. The weakness of *must*: In defense of a mantra. To appear in *Semantics & Linguistic Theory (SALT) 24*.

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

Lewis, David. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic* 8(1).

339–359. doi:10.1007/BF00258436.

Lewis, David. 1981. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic* 10(2). 217–234.

Mellor, D. H. 2005. *Probability: A philosophical introduction*. Routledge.

Narens, Louis. 2007. *Theories of probability: an examination of logical and qualitative foundations*. World Scientific Publishing Company.

Palmer, Frank Robert. 1979. *Modality and the English modals*. Longman London.

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

van Rooij, Robert. 2011. Measurement, and interadjective comparisons. *Journal of Semantics* 28(3). 335–358.

Rullmann, Hotze. 1995. *Maximality in the Semantics of* wh-*constructions*: University of Massachusetts, Amherst dissertation.

Sassoon, Galit. 2010. The degree functions of negative adjectives. *Natural Language Semantics* 18(2). 141–181.

Sauerland, Uli & Penka Stateva. 2007. Scalar vs. epistemic vagueness: Evidence from approximators. In M. Gibson & T. Friedman (eds.), *Proceedings of semantics and Linguistic Theory xvii*, CLC Publications, Cornell University.

Savage, Leonard J. 1954. *The Foundations of Statistics*. Wiley.

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

Stone, Matthew. 1994. The reference argument of epistemic *must*. In *Proceedings of iwcs 1*, 181–190.

Swanson, Eric. 2006. *Interactions With Context*: MIT dissertation. http://mit.dspace.org/bitstream/handle/1721.1/37356/123190034.pdf?sequence=1.

Yalcin, Seth. 2005. A puzzle about epistemic modals. *MIT Working Papers in Linguistics* 51. 231–272.

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

Yalcin, Seth. 2010. Probability Operators. *Philosophy Compass* 5(11). 916–937.

Yalcin, Seth. 2011. Nonfactualism about epistemic modality. In A. Egan & B. Weatherson (eds.), *Epistemic modality*, Oxford University Press.

Zadeh, Lofti A. 1965. Fuzzy sets. *Information and Control* 8(3). 338–353.

Zadeh, Lofti A. 1978. Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems* 1(1). 3–28.

6 Appendix: Proofs

6.1 Proof of (25) (Lassiter 2010a: 210; cf. Yalcin 2010).

(66) a. $\phi \succsim^o \psi$

b. $\phi \succsim^o \chi$

c. $\therefore\ \phi \succsim^o (\psi \vee \chi)$

(66a) is true iff, for every ψ-world v, there is a ϕ-world u such that $u \succcurlyeq_{\mathbf{g}(w)} v$. Likewise, (66b) is true iff, for every χ-world v', there is a ϕ-world u' such that $u' \succcurlyeq_{\mathbf{g}(w)} v'$. Let z be an arbitrary world in $\psi \vee \chi$. Case 1: $z \in \psi$. Then (66a) obviously entails that there is a ϕ-world u such that $u \succcurlyeq_{\mathbf{g}(w)} z$. Case 2: $z \in \chi$. Then (66b) similarly entails that there is a ϕ-world u' such that $u' \succcurlyeq_{\mathbf{g}(w)} z$. Since z was arbitrary, we conclude that for every $z \in (\psi \vee \chi)$, there is a $z' \in \phi$ such that $z' \succcurlyeq_{\mathbf{g}(w)} z$; thus, by the definition of $\succcurlyeq^o$, (25c) holds. □

6.2 Proof that one of the conditions in (31) (=(68)) is true in all countermodels to (30) (=(67)).

(67) a. $\phi \succcurlyeq^n \psi$
b. $\phi \succcurlyeq^n \chi$
c. $\therefore\ \phi \succcurlyeq^n (\psi \vee \chi)$ **(not valid)**

(68) a. $\exists u \in (\psi \wedge \neg\phi)\ \exists v \in (\phi \wedge \chi)[\neg(u \succ_{\mathbf{g}(w)} v)]$
b. $\exists u' \in (\chi \wedge \neg\phi)\ \exists v' \in (\phi \wedge \psi)[\neg(u' \succ_{\mathbf{g}(w)} v')]$

Suppose (67a) and (67b) are true and (67c) is false. Since (67c) is false, there is a world u in $((\psi \vee \chi) \wedge \neg\phi)$ such that, for all v in $(\phi \wedge \neg(\psi \vee \chi))$, $u \succ_{\mathbf{g}(w)} v$. World u is of course either in $(\psi \wedge \neg\phi)$ or in $(\chi \wedge \neg\phi)$. (All world variables in this and the next proof are implicitly restricted to $\bigcap \mathbf{f}(w)$, except for the distinguished actual world variable w.)

Suppose first that $u \in (\psi \wedge \neg\phi)$. Since (67a) holds, there is no world in $(\psi \wedge \neg\phi)$ that strictly dominates all worlds in $(\phi \wedge \neg\psi)$; and so in particular u does not strictly dominate all of these worlds, i.e. there is some world $v' \in (\phi \wedge \neg\psi)$ such that $\neg(u \succ_{\mathbf{g}(w)} v')$. But since u strictly dominates all worlds in $(\phi \wedge \neg(\psi \vee \chi))$, v' must be in $(\phi \wedge \neg\psi \wedge \chi)$. So, of course, v' is in $(\phi \wedge \chi)$. A parallel argument shows that, if u is in $(\chi \wedge \neg\phi)$, there must be some world v'' in $(\phi \wedge \psi)$ such that $\neg(u \succ_{\mathbf{g}(w)} v'')$. But since u is either in $(\psi \wedge \neg\phi)$ or in $(\chi \wedge \neg\phi)$, one of the two conditions in (68) must hold, and in particular there must be worlds either in $(\phi \wedge \psi)$ or in $(\phi \wedge \chi)$. □

6.3 Proof of (32) (=(69)).

(69) a. $(\phi \wedge \neg\psi \wedge \neg\chi) \succcurlyeq^n (\neg\phi \wedge \psi \wedge \neg\chi)$
b. $(\phi \wedge \neg\psi \wedge \neg\chi) \succcurlyeq^n (\neg\phi \wedge \neg\psi \wedge \chi)$
c. $\therefore\ (\phi \wedge \neg\psi \wedge \neg\chi) \succcurlyeq^n ((\neg\phi \wedge \psi \wedge \neg\chi) \vee (\neg\phi \wedge \neg\psi \wedge \chi))$

By the equivalent variant of the definition of $\succcurlyeq^n$ in (29), (69a) means that, for any world $u \in (\neg\phi \wedge \psi \wedge \neg\chi)$ there is a world $v \in (\phi \wedge \neg\psi \wedge \neg\chi)$ such that $\neg(u \succ_{\mathbf{g}(w)} v)$. (Since the propositions compared are disjoint, the restriction to worlds in non-overlapping portions of the propositions compared in the definition of $\succcurlyeq^n$ has no effect.) Similarly, (69b) means that for any world $u' \in (\neg\phi \wedge \neg\psi \wedge \chi)$ there is a world $v' \in (\phi \wedge \neg\psi \wedge \neg\chi)$ such that $\neg(u \succ_{\mathbf{g}(w)} v)$.

Let z be an arbitrary world in $((\neg\phi \wedge \psi \wedge \neg\chi) \vee (\neg\phi \wedge \neg\psi \wedge \chi))$. Case 1: $z \in (\neg\phi \wedge \psi \wedge \neg\chi)$. Then, as we just saw, (69a) entails that there is a world $v \in (\phi \wedge \neg\psi \wedge \neg\chi)$ such that $\neg(z \succ_{\mathbf{g}(w)} v)$, namely v. Case 2: $z \in (\neg\phi \wedge \neg\psi \wedge \chi)$. Then a parallel argument from (69b) shows that there is a world $v' \in (\phi \wedge \neg\psi \wedge \neg\chi)$ such that $\neg(z \succ_{\mathbf{g}(w)} v')$, namely v'. Since z was arbitrary, we conclude that, for any $z \in ((\neg\phi \wedge \psi \wedge \neg\chi) \vee (\neg\phi \wedge \neg\psi \wedge \chi))$, there is some $z' \in (\phi \wedge \neg\psi \wedge \neg\chi)$ such that $\neg(z \succ_{\mathbf{g}(w)} z')$. This is equivalent to $(\phi \wedge \neg\psi \wedge \neg\chi) \succcurlyeq^n ((\neg\phi \wedge \psi \wedge \neg\chi) \vee (\neg\phi \wedge \neg\psi \wedge \chi))$ by the definition of $\succcurlyeq^n$ in (29), and so (69) is valid. □