Probability Modals and Infinite Domains

Adam Marushak

Abstract

Recent years have witnessed a proliferation of attempts to apply the mathematical theory of probability to the semantics of natural language probability talk. These sorts of "probabilistic" semantics are often motivated by their ability to explain intuitions about inferences involving *likely* and *probable*—intuitions that Angelika Kratzer's canonical semantics fails to accommodate through a semantics based solely on an ordering of worlds and a qualitative ranking of propositions. However, recent work by Wesley Holliday and Thomas Icard has been widely thought to undercut this motivation: they present a world-ordering semantics that yields essentially the same logic as probabilistic semantics. In this paper, I argue that the challenge remains: defenders of world-ordering semantics have yet to offer a plausible semantics that captures the logic of comparative likelihood. Holliday & Icard's semantics yields an adequate logic only if models are restricted to Noetherian pre-orders. But I argue that the Noetherian restriction faces problems in cases involving infinitely large domains of epistemic possibilities. As a result, probabilistic semantics remains the better explanation of the data.

1 Introduction

Natural language talk of what is *likely* or what is *probable* has a foot in two worlds: this language would appear to have ties to the mathematical theory of probability, and yet talk of what is likely is commonplace even among speakers with no special mathematical training. How then should we approach the semantics of this language?

Semanticists were initially reluctant to base their theories on the quantitative notion of probability employed by scientists and mathematicians. This latter notion may well have its origins in the folk concept of probability, but it was thought to be a mistake to credit ordinary language users with a tacit grasp of the mature theory itself.[1]

However, there has recently been a striking shift in opinion: semantics based on Kolmogorovian probability (hereafter: *probabilistic semantics*) have become increasingly widespread.[2]

One reason for this change in attitude has been the recognition that non-probabilistic semantics cannot easily account for explicitly quantitative assessments of probability, as in *there is a 70% chance of rain*. But what is perhaps more surprising is that non-probabilistic semantics fail to capture intuitions about even basic inferences involving judgments of comparative likelihood.

For example, on Kratzer's canonical *world-ordering semantics* (Kratzer (1991, 2012)), an ordering on worlds generates an ordering on propositions, which in turn fixes the facts about what is more/less/equally likely than what.[3] But Lassiter (2010, 2011, 2015) and Yalcin (2010) demonstrate that this semantics validates clearly invalid inference patterns like the following:

> The coin's landing heads is as likely as its landing tails. Therefore, the coin's landing heads is as likely as any proposition whatsoever.

In contrast, probabilistic semantics validates a variety of intuitively valid inference patterns and fails to validate obviously invalid inference patterns like the one above.[4]

Nevertheless, recent work by Holliday & Icard (2013a) suggests that the shift away from world-ordering semantics may be premature: they present a world-ordering semantics that yields essentially the same logic as probabilistic semantics. And indeed, their work has been widely thought to demonstrate that world-ordering semantics can capture the logic of comparative likelihood just as well as probabilistic semantics.[5]

[1]See Hamblin (1959, 234), Koopman (1940, 269-270), and Kratzer (2012, 25) for expressions of this sentiment.

[2]Defenders include Cariani (2016), Carr (2015), Lassiter (2010, 2011, 2015, 2017), Moss (2013, 2015, 2018), Rothschild (2012), Swanson (2006, 2011, 2016), and Yalcin (2007, 2010, 2011).

[3]The label *world-ordering semantics* is due to Holliday & Icard (2013b).

[4]See Lassiter (2010, 2011, 2015) and Yalcin (2010).

[5]See Cariani (2016, n. 9), Carr (2015, 697 n. 23), Lassiter (2015, 663), and Suzuki (2013, 216).

However, in this paper, I argue that the challenge remains: defenders of world-ordering semantics have yet to offer a plausible semantics that captures the logic of comparative likelihood. Holliday & Icard's semantics yields an adequate logic only if models are restricted to Noetherian pre-orders: i.e. those on which there is no infinite sequence of distinct worlds $w_1 \preceq w_2 \preceq w_3 \ldots$ where each world in the sequence is ranked at least as high as the preceding world. But I argue that the Noetherian restriction faces problems in cases involving infinitely large domains of epistemic possibilities. As a result, probabilistic semantics remains the better explanation of the data.

2 Probabilistic vs. World-Ordering Semantics

Let's begin by contrasting two approaches to the semantics of natural language probability talk.

Start with a standard propositional language $\mathscr{L}$ enriched with the following operators:

- If ϕ and ψ are sentences of $\mathscr{L}$, then $\ulcorner(\phi \geqslant \psi)\urcorner$ is a sentence of $\mathscr{L}$.
- If ϕ is a sentence of $\mathscr{L}$, then $\ulcorner\Diamond\phi\urcorner$ is a sentence of $\mathscr{L}$.[6]

Sentences of the form $(\phi \geqslant \psi)$ are intended to model natural language judgments of comparative likelihood: ϕ *is at least as likely as* ψ. Sentences of the form $\Diamond\phi$ model talk of epistemic possibility: *it might be that* ϕ.

We also add the following definitions:

- $(\phi > \psi)$ models ϕ *is more likely than* ψ and is defined as $(\phi \geqslant \psi) \wedge \neg(\psi \geqslant \phi)$.
- $\Delta\phi$ models *it is likely that* ϕ and is defined as $(\phi > \neg\phi)$.
- $\Box\phi$ models talk of epistemic necessity—*it must be that* ϕ—and is defined as $\neg\Diamond\neg\phi$.

What, then, is the semantics appropriate to $\mathscr{L}$, in its intended interpretation? There are two main approaches, which diverge in their semantics for $\geqslant$. Before setting out these views, let's give the semantic entries they have in common.

[6] I omit corner quotes from here on out.

Let a model be a tuple $\mathscr{M} = \langle W, S, V\rangle$, where W is a non-empty set (intuitively, a set of metaphysical possibilities), S is a non-empty subset of W (intuitively, an information state or a set of epistemic possibilities), and V is a function assigning elements of $\mathscr{P}(W)$ to atomic sentences $A_{\mathscr{L}}=\{p,q,...\}$ (intuitively, V specifies which proposition is expressed by a given atomic sentence). We then define an interpretation $[\![\cdot]\!]_{\mathscr{M}}$ for $\mathscr{M}$ as follows:

- $[\![\phi]\!]_{\mathscr{M}} = V(\phi)$ if $\phi \in A_{\mathscr{L}}$.
- $[\![\neg\phi]\!]_{\mathscr{M}} = W - [\![\phi]\!]_{\mathscr{M}}$.
- $[\![(\phi \wedge \psi)]\!]_{\mathscr{M}} = [\![\phi]\!]_{\mathscr{M}} \cap [\![\psi]\!]_{\mathscr{M}}$.
- $[\![\Diamond\phi]\!]_{\mathscr{M}} = \{w \in W : [\![\phi]\!]_{\mathscr{M}} \cap S \neq \emptyset\}$.[7]

A sentence ϕ is true at w in $\mathscr{M}$ ($[\![\phi]\!]^{w}_{\mathscr{M}} = 1$) iff $w \in [\![\phi]\!]_{\mathscr{M}}$. A sentence is valid in $\mathscr{M}$ iff it is true at every $w \in W$ in $\mathscr{M}$. A sentence is valid in a class of models $\mathscr{C}$ iff it is valid in every model in the class. A semantics validates an inference pattern in $\mathscr{C}$ iff every model $\mathscr{M}$ in $\mathscr{C}$ and w in $\mathscr{M}$ is such that if the premises of the inference are all true at w in $\mathscr{M}$, then the conclusion of the inference is true at w in $\mathscr{M}$.

Our first approach to the semantics of $\geqslant$ treats judgments of comparative likelihood as qualitative comparisons of propositions, where these comparisons are in turn grounded in a more fundamental ranking of the worlds that comprise each proposition.

For example, on Kratzer's (1991) semantics for $\geqslant$, context delivers a set of propositions O—called the *ordering source*—that induces a pre-order, $\succeq_O$, on the members of S as follows:[8]

$$w \succeq_O w' \text{ iff } \{\alpha \in O : w \in \alpha\} \supseteq \{\alpha \in O : w' \in \alpha\}.$$

Intuitively, the ordering source relevant to $\geqslant$ represents a standard of normality, and worlds are ranked higher the closer they come to matching the normal course of events.

Kratzer then uses this ranking of worlds to determine a ranking, $\succsim$, of propositions:

[7] I follow MacFarlane (2011, 2014) and Yalcin (2007) in using S to determine the interpretation of $\Diamond\phi$, as opposed to an accessibility relation. I opt for the former approach to simplify the semantics, since the question of the world-sensitivity of $\Diamond\phi$ and $\geqslant$ is not at issue in what follows.

[8] A pre-order is a reflexive and transitive binary relation.

$$\alpha \gtrsim \beta \text{ iff } \forall w \in \beta : \exists w' \in \alpha : w' \succeq_O w.$$

That is, α is ranked at least as high as β iff every β-world can be paired with an α-world that is at least as highly ranked.[9] Finally, Kratzer takes $(\phi \geqslant \psi)$ to be true at a world in a model iff $[\![\phi]\!]_{\mathscr{M},S} \gtrsim [\![\psi]\!]_{\mathscr{M},S}$, where $[\![\phi]\!]_{\mathscr{M},S} = [\![\phi]\!]_{\mathscr{M}} \cap S$.[10]

Generalizing from the particulars of Kratzer's approach, *world-ordering semantics* takes models to be the following: $\langle W, S, V, \succeq, \uparrow\rangle$, where $\succeq$ is a pre-order on S, and $\uparrow$ is a lifting operation—that is, a function from $\succeq$ to a binary relation, $\succeq^{\uparrow}$, on $\mathscr{P}(S)$. The role of the lifting operation is to take us from a ranking on worlds to a ranking on propositions. So, for example, Kratzer's definition of $\gtrsim$ is one way to lift a pre-order on worlds to a pre-order on propositions. We finally let $[\![(\phi \geqslant \psi)]\!]^{w}_{\mathscr{M}} = 1$ iff $[\![\phi]\!]_{\mathscr{M},S} \succeq^{\uparrow} [\![\psi]\!]_{\mathscr{M},S}$.[11]

However, as Lassiter (2010, 2011, 2015) and Yalcin (2010) point out, world-ordering semantics with Kratzer's (1991) lifting operation faces what Lassiter (2015) calls the *disjunction puzzle*—namely, the semantics validates the following inference pattern:

I1:
P1. $\phi \geqslant \psi$
P2. $\phi \geqslant \chi$
C. $\phi \geqslant (\psi \vee \chi)$

I1 is clearly invalid: from the fact that heads is at least as likely as heads, and heads is at least as likely as tails, it does not follow that heads is at least as likely as heads or tails.[12]

Lassiter and Yalcin use the disjunction puzzle to motivate an alternative semantics for $\geqslant$. On *probabilistic semantics*, judgments of comparative

[9] This method of generating a ranking of propositions from a ranking of worlds is due to Lewis (1973).

[10] Strictly speaking, Kratzer takes probability talk to have a world-sensitive semantics, but I employ the information state parameter S in order to simplify the various semantic theories discussed in this paper (see n. 7). A referee also notes that in Kratzer (1986), she in fact leaves open the possibility that probability talk has a quantitative, not qualitative semantics.

[11] This generalization of Kratzer's semantics and the term *lifting operation* are due to Holliday & Icard (2013b).

[12] This example is due to Yalcin (2010).

likelihood are grounded in a quantitative ranking of propositions. Models are as follows: $\langle W, S, V, \mathscr{F}, \mu \rangle$, where $\mathscr{F}$ is a σ-algebra of subsets of S, and μ is a finitely additive probability measure. That is, $\mathscr{F}$ is a subset of $\mathscr{P}(S)$ such that $S \in \mathscr{F}$, and $\mathscr{F}$ is closed under complementation and countable union. μ is a function from $\mathscr{F}$ to [0,1] such that $\mu(S) = 1$ and $\mu(\alpha \cup \beta) = \mu(\alpha) + \mu(\beta)$, for disjoint α and β. We then let $[\![(\phi \geqslant \psi)]\!]^w_{\mathscr{M}} = 1$ iff $\mu([\![\phi]\!]_{\mathscr{M},S}) \geq \mu([\![\psi]\!]_{\mathscr{M},S})$. Alternatively, one can give a probabilistic semantics based on a *set* of probability measures P, where $[\![(\phi \geqslant \psi)]\!]^w_{\mathscr{M}} = 1$ iff $\mu([\![\phi]\!]_{\mathscr{M},S}) \geq \mu([\![\psi]\!]_{\mathscr{M},S})$ for every $\mu \in P$.[13]

Probabilistic semantics avoids the disjunction puzzle and also validates a range of intuitively valid inference patterns identified by Yalcin (2010).[14] As a result, there appear to be solid grounds for favoring probabilistic over world-ordering semantics.[15]

3 Holliday & Icard's Alternative

As Kratzer (2012) notes, her (1991) choice of lifting operation is one among many. It thus remains to be seen whether one can formulate an alternative lifting operation that yields better predictions.[16]

Holliday & Icard (2013a,b) claim to do just that: they present an alternative lifting operation that promises to resolve the disjunction puzzle and capture the core inferences involving probability talk:

> *m*-lifting: $[\![\phi]\!]_{\mathscr{M},S} \succeq^m [\![\psi]\!]_{\mathscr{M},S}$ iff there exists an injective function $f : [\![\psi]\!]_{\mathscr{M},S} \to [\![\phi]\!]_{\mathscr{M},S}$ such that $\forall w \in [\![\psi]\!]_{\mathscr{M},S} : f(w) \succeq w$.

[13] See Lassiter (2011, 81) and Rothschild (2012).

[14] See Lassiter (2010, 2011, 2015) and Yalcin (2010).

[15] My focus in this paper is strictly on the debate between probabilistic and world-ordering semantics. Thus, there are numerous options for theorizing about the meaning of natural language probability talk that lie outside the scope of this paper. To mention just a few: I will leave aside the question of whether a probabilistic semantics should be based on finite additivity, countable additivity, or qualitative additivity (see Holliday & Icard (2013b) and Lassiter (2015) for discussion of qualitative additivity; see §5 for discussion of countable additivity). I also leave aside the question of whether probabilistic semantics fares better than what Holliday & Icard (2013b) call *event-ordering semantics*. Finally, it is also worth considering whether probability talk should be understood in terms of ranking functions, plausibility orders, or other tools from the belief revision literature.

[16] Kratzer (2012) proposes a different lifting operation, but Lassiter (2015) demonstrates that the resulting semantics still fails to avoid a version of the disjunction puzzle.

An injective function is one that maps every distinct element in the domain to a distinct element in the codomain: that is, for every w, w' in the domain, if $f(w) = f(w')$, then $w = w'$. Thus, m-lifting tells us that a proposition α is at least as highly ranked as β iff each β-world can be mapped to a distinct α-world that is at least as highly ranked.

It's easy to see how this new lifting operation resolves the disjunction puzzle. Consider an instance of I1: $((p \geqslant q) \wedge (p \geqslant \neg q)) \rightarrow (p \geqslant (q \vee \neg q))$. Suppose $W = S = \{w, w'\}$, where w is the sole p-world and q-world. Our instance of I1 will be false at w if $w \succeq w'$.

Intuitively, m-lifting invalidates I1 because the same p-world can't do double duty in matching up to each of the q and $\neg q$-worlds: this is ruled out by the requirement that each $(q \vee \neg q)$-world be mapped to a *distinct* p-world in order for p to be at least as likely as the disjunction.

World-ordering semantics with m-lifting has further virtues. Not only does the semantics resolve the disjunction puzzle: Harrison-Trainor et al. (2017, 2018) prove that world-ordering semantics with m-lifting yields the exact same logic as the set-of-measures probabilistic semantics when models are restricted to Noetherian pre-orders—i.e. those on which there is no infinite sequence of distinct worlds $w_1 \preceq w_2 \preceq w_3 \ldots$. World-ordering semantics with m-lifting and the Noetherian restriction yields a logic that differs only slightly from that of a single-measure probabilistic semantics. The latter but not the former validates the comparability principle: $(\phi \geqslant \psi) \vee (\psi \geqslant \phi)$.[17]

The upshot would seem to be this: world-ordering semantics captures the logic of comparative likelihood just as well as probabilistic semantics. Consequently, semanticists will have to look elsewhere for grounds favoring one semantics over the other.

To be clear: the threat here is not that there are *no* grounds for favoring one semantics over the other. There are other motivations for probabilistic semantics, and there are questions about whether m-lifting can capture other data about the use of probability modals.[18] But the general lesson of Holliday

[17] See Holliday & Icard (2013b) and Holliday & Icard (2018).

[18] See Lassiter (2010, 2011, 2015, 2017), Moss (2013, 2015, 2018), Rothschild (2012), Swanson (2006, 2011, 2016), and Yalcin (2007, 2011) for discussion of alternative motivations for probabilistic semantics. See Lassiter (2015) for criticisms of Holliday & Icard's semantics distinct from those I raise below. There are further questions regarding how to integrate m-lifting into Kratzer's larger account of modal language. I1 is arguably valid for deontic comparatives, so m-lifting cannot serve as a general lifting operation for all comparative modal language. This means that different flavors of modality vary not just

& Icard (2013a,b) and Harrison-Trainor et al. (2017, 2018) is thought to be the following: whatever the grounds will be for favoring one semantics over the other, they will not concern the logic of comparative likelihood.

4 The Noetherian Restriction

As it turns out, the dialectic is more complex. Holliday & Icard's semantics has two components: (i) m-lifting; (ii) the Noetherian restriction. It is only the conjunction of these two components that delivers an adequate logic for probability talk. But at present, (ii) has received no attention in the literature. I will argue, however, that (ii) faces problems that have no analogue in probabilistic semantics. As a result, defenders of world-order semantics have yet to offer a plausible semantics that captures the logic of comparative likelihood. Instead of leading to a stalemate, the logic of comparative likelihood still supplies a key motivation for probabilistic over world-ordering semantics.

Let's begin by examining the role of the Noetherian restriction. Suppose we simply adopt world-ordering semantics with m-lifting without the Noetherian restriction. The resulting semantics fails to validate an inference pattern that Yalcin (2010) calls V11:

V11:
P1. $\phi \geqslant \psi$
P2. $\Delta\psi$
C. $\Delta\phi$

V11 is clearly a valid inference pattern. For example, if rain is at least as likely as high winds, then if high winds are likely, rain is likely as well.

But the following constitutes a countermodel to V11 for world-ordering semantics with m-lifting and no Noetherian restriction:

Let $\mathscr{M} = \langle W, S, V, \succeq, \uparrow \rangle$, where:

$W = S = \mathbb{N}$

$V(p) = \{x \in \mathbb{N} : x \text{ is even}\}$

in their modal base and ordering source but also in their lifting operation. Thanks to Eric Swanson for discussion on this point.

$\succeq$ is a flat ranking: for every $w, w' \in S, w \succeq w'$

$\uparrow = m$-lifting

Choose any $w \in W$. The following instance of V11 is false at w:

(*) $((p \geqslant (p \vee \neg p)) \wedge \Delta(p \vee \neg p)) \rightarrow \Delta p$

To see why, notice that since every pair of worlds in S is equally ranked, establishing the first conjunct of the antecedent of (*) simply requires showing that there exists an injection from the $(p \vee \neg p)$-worlds to the p-worlds. Here is such an injection: $f(x) = 2x$. Next, note that the second conjunct of the antecedent is true at w since there is trivially an injection from the empty set to $\mathbb{N}$, but not vice versa. However, the consequent of the conditional is false at w since $(\neg p \geqslant p)$ is true at w: $f(x) = (x + 1)$ is an injection from the p-worlds to the $\neg p$-worlds.[19]

What drives this countermodel is the following distinctive feature of infinitely large sets: a proper subset of a set may have the same cardinality as the set itself. In this case, $V(p)$ is a subset of $\mathbb{N}$ that has the same cardinality as $\mathbb{N}$. Hence, since each element of $\mathbb{N}$ is ranked equally high, p will be at least as likely as $\mathbb{N}$.[20] But the entire domain, $\mathbb{N}$, is trivially likely, while p is not, since its complement is equally likely. Hence, the semantics fails to validate V11.[21]

[19] Our model also constitutes a countermodel to what Holliday & Icard call V13:

V13:
P1. $(\phi \wedge \neg\psi) > \bot$
C. $(\phi \vee \psi) > \psi$

The following instance is false at any $w \in W$: $(((p \vee \neg p) \wedge \neg p) > \bot) \rightarrow (((p \vee \neg p) \vee p) > p)$. See §6 for further discussion of V13. Note also that our model serves as a countermodel to V11 and V13 even if $\Delta\phi$ is stronger than $(\phi > \neg\phi)$—the countermodel requires only that $\Delta\phi$ entail $(\phi > \neg\phi)$.

[20] This itself is an odd result and suggests that the semantics still does not handle disjunctions properly.

[21] We can construct a similar countermodel with an uncountably infinite domain. Let $W = S = [0, 2\pi)$, let $V(p) = [0, \pi]$, and retain the rest of our original countermodel. (*) is false in this model for the same reason: the cardinality of $V(p)$ is the same as that of S and that of $[\![\neg p]\!]$.

The same countermodel also shows that world-ordering semantics with m-lifting and no Noetherian restriction fails to validate the following inference pattern, which I'll call V14:

V14:
P1. $\phi \geqslant \top$
C. $\Delta\phi$

By contrast, it's trivial to verify that probabilistic semantics validates V11 and V14 regardless of the size of S.

These results are clearly problematic for world-ordering semantics with m-lifting and no Noetherian restriction: V11 and V14 are intuitively valid regardless of the size of the domain of epistemic possibilities. For example, suppose you learn that there is a finite number of stars, but you do not know the finite upper bound. If you then learn that there being an even number of stars is at least as likely as a tautology, you should obviously conclude that it is likely that there is an even number of stars.[22] However, on world-ordering semantics with m-lifting and no Noetherian restriction, this conclusion does not follow.

The upshot: the Noetherian restriction plays a crucial role in Holliday & Icard's semantics. If we remove this restriction, holding fixed the other elements of their semantics, their theory invalidates obviously valid inference patterns.[23]

[22]This interpretation of the countermodel is inspired by an example from Portner (2009, 33).

[23]I leave open the possibility that Holliday & Icard might supplement their theory in other ways in order to secure the correct logic through m-lifting without relying on the Noetherian restriction. One option, of course, is simply to restrict models to those with finite domains. One might instead appeal to differences in density between subsets of an infinite set as compared with the entire set itself. But these options lead to other problems, e.g. infinite domains are plausibly necessary for representing natural language meaning, and appealing to density arguably relies on quantitative probability. A referee suggests another strategy: one might allow for non-Noetherian models but hold that $\succeq^{\uparrow}$ is defined only if the cardinality of the compared propositions is finite. However, this strategy rules out the possibility of capturing intuitively true comparisons such as: there being an even number of stars is more likely than the epistemically impossible scenario in which there are no stars.

5 Problems with the Noetherian Restriction

I will present two problems for the Noetherian restriction. But let me first raise an objection to Holliday & Icard's grounds for making the restriction in the first place.

Holliday & Icard motivate the Noetherian restriction by analogy with a similar constraint on probabilistic semantics:

> [W]e assume world-ordering models are *Noetherian*: there is no infinite sequence $w_1 \preceq w_2 \preceq w_3 \ldots$ of distinct worlds, just as with a finitely additive measure, there is no infinite sequence of distinct worlds with non-zero, non-decreasing measure (Holliday & Icard (2013a, 526)).[24]

But this analogy is flawed. One cannot have a finitely additive, non-zero, non-decreasing measure over an infinite domain—but one *can* have a finitely additive measure over an infinite domain that assigns probability zero to each outcome. Indeed, one of the principal motivations for finite but not countable additivity is that it allows one to assign probability zero to each outcome in a countably infinite domain and thereby capture the judgment that each outcome is equally likely.[25] Furthermore, if the sample space is continuous—i.e. if it contains an uncountably infinite set of possible outcomes—then we *must* assign probability zero to an uncountable number of outcomes.[26]

Defenders of probabilistic semantics are aware of these facts. Several of them explicitly allow for assigning probability zero to every outcome in an infinite domain.[27] Thus, without further argument, there is no reason to believe that defenders of probabilistic semantics have to make a stipulation analogous to the Noetherian restriction. The former theorists can allow for assigning probability zero to every outcome in an infinite domain, but Holliday & Icard cannot allow for a flat ranking over an infinite domain: as we saw in §4, models with such rankings invalidate V11 and V14.

Now, the first problem with the Noetherian restriction itself concerns the origin of the ranking on worlds. The standard account—due to Kratzer (1981, 1991, 2012)—is that the ranking on worlds is fixed by a contextually-determined set of ordering source propositions, as we reviewed in §2. But

[24]Holliday & Icard (2018, 87) offers the same motivation for the Noetherian restriction.
[25]See de Finetti (1974).
[26]See Williamson (2007).
[27]See Cariani (2016), Carr (2015), and Yalcin (2007).

suppose the set of ordering source propositions is empty: every world will trivially verify the same ordering source propositions and will thus be equally ranked. Consequently, if the domain of epistemic possibilities is infinite, the Noetherian restriction precludes an empty ordering source (recall that a flat ranking on an infinite domain is not a Noetherian model).

This is a bad result. Epistemic possibility modals are often thought to have readings that involve an empty ordering source, so why should the situation be any different for probability modals, or different when the domain is infinite?[28] Furthermore, the ordering source for probability modals is supposed to represent a contextually determined standard of normality. But the context of use might simply fail to determine some such standard and thereby leave the ordering source empty. The Noetherian restriction thus conflicts with the standard account of what determines the ranking on worlds in the first place: the mere fact that the domain is infinite should not rule out the possibility of an empty ordering source.

The second problem with the Noetherian restriction concerns our intuitions about examples involving infinite domains. Recall the star interpretation of our countermodel discussed above: suppose we are wondering how many stars there are in the universe, and we are unwilling to place a finite upper bound on the answer. Here our information state is best modeled by a countably infinite domain of possible worlds, each containing an ever-greater number of stars. Now, *prima facie*, it is possible to be in a coherent information state of this sort according to which, above a certain threshold, any number of stars is as likely as any other number of stars. Or suppose we are wondering about the precise value of some physical constant. Again, it is plausible that there exists some state of belief or evidence according to which each of an uncountably infinite range of values is equally likely.[29] One candidate for such an information state is that of total ignorance.[30] There are

[28]See Faller (2011) and Peterson (2010) for inventive applications of empty-ordering-source readings of epistemic possibility modals. Note that Kratzer herself allows for such readings (see Kratzer (1981)).

[29]Cf. Easwaran (2014, 19–20).

[30]It is controversial whether a single probability measure can adequately represent a state of total ignorance. A referee notes that there is no constructive proof of a uniform distribution over the natural numbers (see Lauwers (2009)). And the non-constructive choices required to generate such a distribution seem at odds with the distribution's modeling a state of total ignorance. However, the set-of-measures semantics discussed at the end of §2 may help avoid this problem. Also, it may be easier to model ignorance when the domain is uncountably infinite—e.g. consider the Lebesgue measure on [0,1], on which

others. One might possess—or merely *believe* that one possesses—positive evidence that the value of this constant is determined by a random process. All of this is to say: our semantics for probability modals should not rule out the possibility that each of an infinite set of outcomes is equally likely.

But the Noetherian restriction does rule this out. Equiprobability of outcomes requires a flat ranking—one that is not Noetherian if the domain of epistemic possibilities is infinitely large.[31] By contrast, probabilistic semantics allows for equiprobability of outcomes across an infinite domain. One need only assign each outcome probability zero.[32]

To sum up: the Noetherian restriction secures an adequate logic at the cost of (a) introducing an unmotivated constraint on world-orders; (b) precluding an empty ordering source when the domain is infinite; (c) ruling out the possibility of equiprobable outcomes across an infinite domain.

6 Partners in Crime?

In defending probabilistic semantics, I've appealed to probability-zero epistemic possibilities. But one might wonder whether allowing such possibilities leads to undesirable consequences. If so, one could argue that a viable probabilistic semantics must indeed make a stipulation analogous to the Noetherian restriction: models cannot include a measure μ such that $\mu(\{w_1\}) \leq \mu(\{w_2\}) \leq \mu(\{w_3\}) \ldots$ for an infinite sequence of distinct worlds. Such a measure is possible only if each world is assigned probability zero (any greater value would violate the requirement that $\mu(S) = 1$).[33]

---

the probability of every interval in [0,1] is equal to its length.

[31] A world-order in which each possibility is incomparable will deliver the result that none of the outcomes are more or less likely than the others. However, incomparability of outcomes is not the same as equiprobability—i.e. that each is equally likely.

[32] Equiprobability of outcomes across a countably infinite domain requires a finitely but not countably additive measure. Equiprobability across an uncountably infinite domain is consistent with countable additivity.

[33] Lassiter (2015, 2016, 2017) avoids relying on probability-zero epistemic possibilities by appealing to the non-zero granularity of natural language meaning. He argues that sentences that appear to describe real number quantities with probability zero (e.g. that a car is going exactly 35 miles per hour) really express propositions about a range of values (e.g. that the car is going 35 miles per hour plus or minus some non-zero, real number g), where the range has non-zero probability. But as a referee notes, Lassiter's strategy cannot accommodate intuitions of equiprobability across a countably infinite domain, since here our intuitions concern precise natural number quantities that must all receive probability

So: what sort of undesirable consequences might follow from allowing probability-zero possibilities? It is true that if we allow for probability-zero possibilities, leaving the rest of our probabilistic semantics unchanged, our theory will invalidate a plausible principle connecting epistemic possibility and comparative likelihood:

$\textit{Regularity}_C$: $\Diamond\phi \rightarrow (\phi > \bot)$[34]

Recall that on the probabilistic semantics discussed in §2, facts about comparative likelihood are settled by comparing the measure value of each proposition. Thus, since contradictions receive probability zero, no probability-zero possibility will be more likely than a contradiction. This is a highly unintuitive result. Returning to our star case above, it seems absurd to claim that a googol stars in the universe is no more likely than $0 = 1$.[35]

However, it is possible to amend our probabilistic semantics to validate $\textit{Regularity}_C$ while allowing for probability-zero possibilities. To see how, first distinguish $\textit{Regularity}_C$ from a similar principle, also called "Regularity", that is often discussed in formal epistemology and probability theory:

$\textit{Regularity}_P$: if $\alpha \neq \emptyset$ and $\alpha \in \mathscr{F}$, then $\mu(\alpha) > 0$.[36]

$\textit{Regularity}_P$ simply expresses a constraint on probability measures, yet it's natural to think that $\textit{Regularity}_P$ is the only way to secure $\textit{Regularity}_C$ in a probabilistic semantics. If so, one cannot validate $\textit{Regularity}_C$ if one accepts probability-zero possibilities: such possibilities violate $\textit{Regularity}_P$, since they would be non-empty members of $\mathscr{F}$ that do not receive greater-than-zero probability.

But the natural thought is false. There are other routes to validating $\textit{Regularity}_C$ in a probabilistic semantics. Easwaran (2014, 16) captures the general idea:

> What we need is some mathematical relation $p \succ q$ that says when p is more likely than q. But this relation can depend on mathematical facts beyond $P(p)$ and $P(q)$. ... [S]tandard probabilism gives two

zero (recall the star example from §5).

[34]The subscript C indicates that the principle constrains judgments of comparative likelihood; I discuss a different type of regularity principle below.

[35]This case is adapted from an example due to [reference removed for blind review].

[36]"Regularity" is sometimes formulated as the stronger thesis that all non-empty subsets of S receive positive probability.

further mathematical features that might be relevant—the conditional probability function $P(\cdot|\cdot)$, and the set $[S]$ of doxastic possibilities.

Easwaran (2014, 16–17) goes on to offer several proposals for maintaining *Regularity*$_C$ without *Regularity*$_P$ by using Popper's (1955) axioms for conditional probability. I offer another alternative here, which exploits features of the set S of doxastic possibilities:[37]

Modified Probabilistic Semantics:

$$\begin{aligned}[\![(\phi \geqslant \psi)]\!]^w_{\mathscr{M}} = 1 \text{ iff } \quad &\text{(i) } \mu([\![\phi]\!]_{\mathscr{M},S}) \geq \mu([\![\psi]\!]_{\mathscr{M},S}),\\ &\text{(ii) } [\![(\Diamond\psi \to \Diamond\phi)]\!]^w_{\mathscr{M}} = 1, \text{ and}\\ &\text{(iii) } [\![(\Box\psi \to \Box\phi)]\!]^w_{\mathscr{M}} = 1.\end{aligned}$$

On this semantics, judgments of comparative likelihood are sensitive to the epistemic possibility or necessity of the propositions so compared—not just their measure values. That is, ϕ's being at least as likely as ψ requires not just an equal or higher measure value; ϕ must also match up to ψ qualitatively, being epistemically possible if ψ is, and epistemically necessary if ψ is. These qualitative comparisons are responsible for yielding the correct predictions for comparisons involving probability zero and probability one propositions. In particular, clause (ii) ensures that a contradiction is never at least as likely as an epistemically possible proposition—even if the latter has probability zero. But any proposition is at least as likely as a contradiction, so an epistemically possible proposition will always be more likely than a contradiction. Clause (iii) ensures that epistemically necessary propositions will be more likely than those that merely receive probability one. This is a desirable result: additivity requires that the negation of any probability-zero possibility receive probability one, but such propositions—e.g. the proposition that the number of stars is not equal to a googol—are clearly less likely than epistemic necessities, such as the proposition that the number of stars is either even or odd. Finally, the semantics validates V11, V14, and the other validities on Yalcin's (2010) list.

[37] I leave it as an open question which probabilistic semantics is best for maintaining *Regularity*$_C$ without *Regularity*$_P$. Detailed comparison of the various alternatives is outside the scope of this paper. My goal in this section is simply to show that abandoning *Regularity*$_P$ is a plausible move for defenders of probabilistic semantics.

Now, the *Modified Probabilistic Semantics* will fail to validate what Holliday & Icard call V13:

V13:
P1. $(\phi \wedge \neg\psi) > \bot$
C. $(\phi \vee \psi) > \psi$

But it's not obvious that this inference is actually valid. I take it that V13 is not self-evident in the way that V11 or V14 is. Rather, V13 reflects something like the following line of reasoning: if $(\phi \wedge \neg\psi)$ is more likely than a contradiction, then it's possible for ϕ to occur without ψ; but then there are more ways for $(\phi \vee \psi)$ to be true than for ψ itself to be true—namely, all of the ψ-ways plus the $(\phi \wedge \neg\psi)$-ways. This is a valid line of reasoning when the domain of possibilities is finite. Not so when the domain is infinite. Suppose ϕ expresses the proposition that the number of stars is a multiple of 3, and ψ expresses the proposition that the number of stars is a multiple of 6. It is possible for the number of stars to be a multiple of 3 and not a multiple of 6, but the cardinality of $\{w :$ the number of stars in w is a multiple of 3 or a multiple of 6$\}$ is the same as the cardinality of $\{w :$ the number of stars in w is a multiple of 6$\}$. Thus, there is a perfectly respectable notion of size according to which the above line of reasoning goes wrong when the domain is infinite. Of course, others might wish to understand size in terms of the proper superset relation, in which case there is nothing wrong with the above line of reasoning even in the case of an infinite domain. But this only shows that we reach a standoff over V13.[38]

Still, it might be thought that the *Modified Probabilistic Semantics* falters on a related point. Surely there being 10^{10} *or* 10^{11} stars is more likely than there being 10^{11} stars, but if each disjunct is assigned probability zero, the disjunction will be just as likely as either disjunct.[39] Again though, infinite domains reveal a problem. Intuitively, the disjunction is more likely because there are more ways for it to be true. But it's easy to miss that when the domain is infinite, the disjunction and each disjunct are still *false* in the same number of worlds—namely, a countably infinite number. We can then reason as follows. The disjunction and each disjunct are just as likely to be false, since they are false in the same number of worlds and there are no grounds

[38] Cf. McCall & Armstrong (1989).
[39] Cf. Pruss (2014).

for thinking that the set of worlds in which one is false is more likely than the set of worlds in which the other is false. And since the disjunction and each disjunct are equally likely to be false, they are equally likely to be true.

What this shows is that we confront a puzzle. Let $V(p_{10}) = \{w :$ there are 10^{10} stars in $w\}$, let $V(p_{11}) = \{w :$ there are 10^{11} stars in $w\}$, and model judgments of equal likelihood as follows: $\phi \triangleq \psi =_{df} ((\phi \geqslant \psi) \wedge (\psi \geqslant \phi))$. Each of the following is plausible, but they are jointly inconsistent:

(a) $(p_{10} \vee p_{11}) > p_{11}$

(b) $\neg(p_{10} \vee p_{11}) \triangleq \neg p_{11}$

(c) If $\neg\phi \triangleq \neg\psi$, then $\phi \triangleq \psi$.

I suggest we reject (a). Its plausibility results from the failure to recognize that when the domain is infinite, each sentence is false in the same number of worlds. So our semantics indeed yields the correct verdict about the case.

I take the upshot to be the following. Probabilistic semantics does not bear the same costs as Holliday & Icard's world-ordering semantics with the Noetherian restriction: there exists a well-motivated probabilistic semantics that captures the logic of comparative likelihood, allows for equiprobable outcomes across an infinite domain, and retains *Regularity*$_C$. As a result, probabilistic semantics still provides a better account of the inference patterns governing natural language probability talk.

To be sure, nothing I've said rules out the possibility that an alternative lifting operation can capture the logic of comparative likelihood without the Noetherian restriction. But it remains an open question whether such a lifting operation exists. For now, then, the data favors probabilistic semantics over world-ordering semantics.

References

F. Cariani (2016). 'Deontic Modals and Probability: One Theory to Rule Them All?'. In N. Charlow & M. Chrisman (eds.), *Deontic Modality*. Oxford University Press.

J. Carr (2015). 'Subjective Ought'. *Ergo, an Open Access Journal of Philosophy* **2**.

B. de Finetti (1974). *Theory of Probability, Vols 1 and 2*. New York: John Wiley.

K. Easwaran (2014). 'Regularity and Hyperreal Credences'. *Philosophical Review* **123**(1):1–41.

M. Faller (2011). 'A Possible Worlds Semantics for Cuzco Quecha Evidentials'. In *Proceedings of SALT 20*.

C. L. Hamblin (1959). 'The Modal "Probably"'. *Mind* **68**(270):234–240.

M. Harrison-Trainor, et al. (2017). 'Preferential Structures for Comparative Probabilistic Reasoning'. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*.

M. Harrison-Trainor, et al. (2018). 'Inferring Probability Comparisons'. *Mathematical Social Sciences* **91**:62–70.

W. H. Holliday & T. Icard (2013a). 'Measure Semantics and Qualitative Semantics for Epistemic Modals'. Postprint of publication in Proceedings of SALT 23; URL: `https://philosophy.berkeley.edu/file/869/Holliday_and_Icard_SALT_23.pdf`.

W. H. Holliday & T. Icard (2018). 'Axiomatization in the Meaning Sciences'. In D. Ball & B. Rabern (eds.), *The Science of Meaning: Essays on the Metatheory of Natural Language Semantics*, pp. 73–97. Oxford, UK: Oxford University Press.

W. H. Holliday & T. F. Icard (2013b). 'Measure Semantics and Qualitative Semantics for Epistemic Modals'. *Proceedings of SALT 23* pp. 514–534.

B. O. Koopman (1940). 'The Axioms and Algebra of Intuitive Probability'. *Annals of Mathematics* pp. 269–292.

A. Kratzer (1981). 'The Notional Category of Modality'. In Eikmeyer & Rieser (eds.), *Words, Worlds, and Contexts, new approaches to word semantics*, pp. 38–74. Walter de Gruyter, Berlin.

A. Kratzer (1986). 'Conditionals'. *Chicago Linguistics Society* **22**(2):1–15.

A. Kratzer (1991). ‘Modality’. In A. von Stechow & D. Wunderlich (eds.), *Semantics: An International Handbook of Contemporary Research*, pp. 539–50. Walter de Gruyter, Berlin.

A. Kratzer (2012). *Modals and Conditionals: New and Revised Perspectives*. Oxford University Press.

D. Lassiter (2010). ‘Gradable Epistemic Modals, Probability, and Scale Structure’. In *Proceedings of SALT 20*.

D. Lassiter (2011). *Measurement and Modality: The Scalar Basis of Modal Semantics*. Ph.D. thesis, New York University.

D. Lassiter (2015). ‘Epistemic Comparison, Models of Uncertainty, and the Disjunction Puzzle’. *Journal of Semantics* **32**(4):649–684.

D. Lassiter (2016). ‘Linguistic and Philosophical Considerations on Bayesian Semantics’. In M. Chrisman & N. Charlow (eds.), *Deontic Modals*, pp. 82–116. Oxford, UK: Oxford University Press.

D. Lassiter (2017). *Graded Modality: Qualitative and Quantitative Perspectives*. Oxford, UK: Oxford University Press.

L. Lauwers (2009). ‘The Uniform Distributions Puzzle’. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1393712.

D. K. Lewis (1973). *Counterfactuals*. Blackwell.

J. MacFarlane (2011). ‘Epistemic Modals are Assessment Sensitive’. In A. Egan & B. Weatherson (eds.), *Epistemic Modality*. Oxford University Press.

J. MacFarlane (2014). *Assessment Sensitivity*. Oxford University Press, Oxford.

S. McCall & D. M. Armstrong (1989). ‘God’s Lottery’. *Analysis* **49**(4):223–224.

S. Moss (2013). ‘Epistemology Formalized’. *Philosophical Review* **122**(1):1–43.

S. Moss (2015). 'On the Semantics and Pragmatics of Epistemic Vocabulary'. *Semantics and Pragmatics* **8**(5):1–81.

S. Moss (2018). *Probabilistic Knowledge*. Oxford, UK: Oxford University Press.

T. Peterson (2010). *Epistemic Modality and Evidentiality in Gitksan at the Semantics-Pragmatics Interface*. Ph.D. thesis, The University of British Columbia.

K. R. Popper (1955). 'Two Autonomous Axiom Systems for the Calculus of Probabilities'. *British Journal for the Philosophy of Science* **6**(21):51–57.

P. Portner (2009). *Modality*. Oxford University Press, Oxford.

A. R. Pruss (2014). 'Infinitesimals Are Too Small for Countably Infinite Fair Lotteries'. *Synthese* **191**(6):1051–1057.

D. Rothschild (2012). 'Expressing Credences'. *Proceedings of the Aristotelian Society* **112**(1pt.1):99–114.

S. Suzuki (2013). 'Epistemic Modals, Qualitative Probability, and Nonstandard Probability'. In *Proceedings of the 19th Amsterdam Colloquium*.

E. Swanson (2006). *Interactions with Context*. Ph.D. thesis, Massachusetts Institute of Technology.

E. Swanson (2011). 'How Not to Theorize about the Language of Subjective Uncertainty'. In A. Egan & B. Weatherson (eds.), *Epistemic Modality*. Oxford University Press.

E. Swanson (2016). 'The Application of Constraint Semantics to the Language of Subjective Uncertainty'. *Journal of Philosophical Logic* **45**(2):121–146.

T. Williamson (2007). 'How Probable is an Infinite Sequence of Heads?'. *Analysis* **67**(295):173–180.

S. Yalcin (2007). 'Epistemic Modals'. *Mind* **116**(464):983–1026.

S. Yalcin (2010). 'Probability Operators'. *Philosophy Compass* **5/11**:916–937.

S. Yalcin (2011). 'Nonfactualism about Epistemic Modality'. In A. Egan & B. Weatherson (eds.), *Epistemic Modality*. Oxford University Press.