THE REVIEW OF SYMBOLIC LOGIC
Volume 8, Number 4, December 2015

HOW TO EXPRESS SELF-REFERENTIAL PROBABILITY. A KRIPKEAN PROPOSAL

CATRIN CAMPBELL-MOORE
Munich Center for Mathematical Philosophy, Ludwig-Maximilians-Universität München

Abstract. We present a semantics for a language that includes sentences that can talk about their own probabilities. This semantics applies a fixed point construction to possible world style structures. One feature of the construction is that some sentences only have their probability given as a range of values. We develop a corresponding axiomatic theory and show by a canonical model construction that it is complete in the presence of the ω-rule. By considering this semantics we argue that principles such as introspection, which lead to paradoxical contradictions if naively formulated, should be expressed by using a truth predicate to do the job of quotation and disquotation and observe that in the case of introspection the principle is then consistent.

§1. Introduction. We are interested in languages that include sentences that can talk about their own probabilities. In such languages contradictions can arise between seemingly harmless principles such as probabilism and introspection. The sentence that is used to display this is:

(π) The probability of π is less than $1/2$.

Caie (2013) has recently used this as a prima facie argument against probabilism. A possible (but, we will argue, wrong) response to this is to argue that the contradiction is only due to the self-referential nature of the sentence so should not be worried about. A natural way to account for that intuition is to prevent such sentences from appearing in the language. However, we shall argue that the result of doing that is that one cannot properly represent quantification or formalise many natural language assertions, so we think that that is the wrong path to take. Instead we will suggest that such self-referential probability assertions should be expressible, but one should work out how to deal with this language and how to circumvent such contradictions. This is what we will do in this paper. The language we will work with formalises the probability notion as a predicate that applies to the codes of sentences and rational numbers; we will have a sentence like "$\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$" whose intended interpretation is "The probability of φ is $\geqslant \alpha$".

We believe such considerations will become relevant in disciplines that use probabilistic methods, such as formal epistemology and philosophy of science, as these disciplines start to work with formal languages that are more expressive.

The paper is structured as follows. In Section 2 we will give the aforementioned argument that one *should* consider languages that are able to express self-referential probabilities. In that section we will also discuss previous work on self-referential probabilities to put this paper in context.

Received: September 23, 2014.

© Association for Symbolic Logic, 2015
doi:10.1017/S1755020315000118

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

In Section 3, we will motivate and present our suggested semantics, which generalises a very influential theory of truth that originates in a paper by Saul Kripke (1975). Kripke's theory of truth was developed to account for the liar sentence, which is a sentence that says of itself that it is not true. In his paper Kripke constructs an extension of the truth predicate by formalising the procedure of evaluating a sentence. He uses three-valued logics to build up this extension, but the extension of the truth predicate can then be used within classical logic to give a classical model of the language with a truth predicate. In this semantics one will have that for some sentences, such as the liar sentence, $\neg \mathsf{T}^{\ulcorner}\varphi^{\urcorner}$ and $\neg \mathsf{T}^{\ulcorner}\neg\varphi^{\urcorner}$ are both satisfied. In this paper we shall present a generalisation of this semantics to also account for probability predicates. The final semantics we propose will be classical, but for some sentences we will only assign a range of probability values not a particular probability value. So we might have $\neg \mathsf{P}_{>}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}0^{\urcorner})$ and $\neg \mathsf{P}_{<}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}1^{\urcorner})$, but only $\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}0^{\urcorner})$ and $\mathsf{P}_{\leqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}1^{\urcorner})$. Our generalisation follows ideas from Halbach & Welch (2009) where Halbach and Welch develop a semantics for necessity, conceived of as a predicate, by applying Kripke's construction to "possible world" structures in the form of Kripke models from modal logic. We will use probabilistic modal structures to provide the background structure for our construction. This therefore allows one to use the technical advantages of these structures, which might have been thought to only be available when the probability notion is conceived of as an operator (see Halbach *et al.*, 2003).

In Section 4 we give some observations regarding the developed semantics. In Stern (2014a, 2014b), Stern argues that when stating principles about necessity, the job of quotation and disquotation should be done by a truth predicate. We argue for the same thing here: we argue that principles such as introspection are properly expressed by using the truth predicate. In our language the (positive) introspection principle will then be written as:

$$\mathsf{T}^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner} \rightarrow \mathsf{P}_{=}(^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})$$

This allows one to avoid inconsistency and is well-motivated in this semantic construction. In Section 4 we also consider σ-additive probabilities and show that if the underlying probabilistic modal structure has σ-additive probability measures, then the resulting semantics will satisfy the version of the Gaifman condition that is appropriate in our framework. The Gaifman condition is the requirement that

$$P(\exists x \varphi(x)) = \lim_{n \to \infty} P\left(\varphi(\overline{0}) \vee \varphi(\overline{1}) \vee \ldots \vee \varphi(\overline{n})\right).$$

This is interesting because the Gaifman condition has proved challenging in previous work on self-referential probabilities (see Section 2).

In Section 5 we shall give an axiomatic theory that is intended to capture the semantics. Such a theory is important because it allows one to reason about the semantics. As was discussed in Aumann (1999), when one gives a possible worlds framework to formalise a game theory context the question arises of what the players know about the framework itself and this question is best answered by providing a corresponding syntactic approach. Our theory complete in the presence of the ω-rule, which allows one to conclude $\forall x \varphi(x)$ from all the instances of $\varphi(\overline{n})$. This is needed to fix the standard model of arithmetic. To show the completeness when the ω-rule is present we construct a canonical model. This axiomatisation is substantially new research.

Finally, we finish the paper with some conclusions in Section 6.

§2. Self-referential probabilities. So why should such self-referential probability sentences be expressible? Probability is a very useful concept and it is interesting to study

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

languages that have at least a minimal ability to talk about probabilities. We therefore consider languages that can at least formalise expressions such as:

The probability of the coin landing heads is $1/2$.

We also want to be able to express embeddings of probabilities, as this is useful to express agents beliefs about other agents' beliefs, or generally relationships between different notions of probability. We will work with frameworks that assign probabilities to *sentences* instead of *events*, which are subsets of a sample space. Although this is uncommon in mathematical study of probability it is common in logical and philosophical work and it will allow us to give the syntax of our language that can deal with embeddings of probabilities, without first developing a semantics. We will then have constructions such as

$$\mathsf{P}^{\mathrm{A}}(\ldots \mathsf{P}^{\mathrm{B}}(\ldots \mathsf{P}^{\mathrm{A}}\ldots)\ldots)\ldots$$

We furthermore allow for self-applied probability notions, or higher order probabilities, namely constructions such as

$$\mathsf{P}^{\mathrm{A}}(\ldots \mathsf{P}^{\mathrm{A}}\ldots)\ldots$$

These offer us two advantages. Firstly, they allow for a systematic syntax once one wishes to allow for embedded probabilities. Secondly, their inclusion may be fruitful, as was argued for in Skyrms (1980). For example, we can then represent introspective abilities of an agent, or the uncertainty or vagueness about the first order probabilities. If one disagrees and wishes to argue that they are trivial and collapse to the first level, then one should still not prevent them being expressed in the language, but should instead include an extra principle to state this triviality of the higher levels, such as adding an introspection principle, which is something formalising:

If the probability of φ is $\geqslant \alpha$,
then the probability of "The probability of φ is $\geqslant \alpha$" is 1.

In fact, once we have languages that can express self-referential probabilities we see that this introspection principle, along with the analogous negative introspection principle, cannot be satisfied,[1] suggesting that the triviality of the higher levels of probability is more substantial an assumption than it seems at first sight.

There are two ways of giving languages that can express higher order probabilities that do not allow for self-referential probabilities. The first is to consider a hierarchy of languages. This is given by a language L_0 that cannot talk about probabilities at all, together with a metalanguage L_1 that can talk about probabilities of the sentences of L_0, together with another metalanguage L_2 that can talk about the probabilities of sentences of L_1 and L_0 etc. This leads to a sequence of language $L_0, L_1, L_2, \ldots$ each talking about probabilities of the previous languages. In ordinary language we can talk about multiple probability notions, such as objective chance and the degrees of beliefs of different agents, but the different notions should be able to apply to all the sentences of our language and not have a hierarchy of objective chance notions $\mathsf{ch}_0, \mathsf{ch}_1, \ldots$ applying to the different levels of language.

The second approach is to instead consider one language where the probability notion is formalised by an operator. This is the approach taken in Aumann (1999), Fagin *et al.* (1990), Ognjanović & Rašković (1996) and Bacchus (1990), amongst others. Each of these differ in their exact set-up but the idea is that one adds a recursive clause saying: if φ is

[1] If the probability notion satisfies the axioms of probability.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

a sentence of the language then we can form another sentence of the language that talks about the probability of φ. For example in Aumann (1999) and Ognjanović & Rašković (1996) one adds the clause

$$\text{If } \varphi \in L \text{ then } \mathrm{P}_{\geqslant \alpha}(\varphi) \in L$$

to the construction rules of the language L.[2] In this language $\mathrm{P}_{\geqslant \alpha}$ acts syntactically like $\neg$ instead of like a predicate so this is not a language of first order logic.

Both the typed and operator languages do not allow for self-referential probabilities but they also cannot easily account for quantification over all of the sentences of the language. So for example they cannot express:

Annie is certain that Billy has some non-extremal degrees of belief.

There is an alternative language for reasoning about probabilities that can express this quantification: one can add the probability notion as a standard predicate symbol or function symbol in first order logic. This is the approach we will take. In our language, we will be able to formalise the above by:

$$\mathsf{P}^{\mathrm{A}}_{=}\left(\ulcorner \exists x(\mathsf{P}^{\mathrm{B}}_{>}(x, \ulcorner 0 \urcorner) \wedge \mathsf{P}^{\mathrm{B}}_{<}(x, \ulcorner 1 \urcorner))\urcorner, \ulcorner 1 \urcorner\right)$$

If one takes Peano arithmetic as a background theory, then one can derive the diagonal lemma for this language and therefore result in admitting sentences that talk about their own probabilities. Such self-referential probabilities therefore arise when we consider languages that can express such quantification.

Working with languages that can express self-referential probabilities can also be an advantage. In natural language we can assert sentences that are self-referential or not depending on the empirical situation, and an appropriate formal language representing natural language should be able to do this too.[3]

Consider the following example. Suppose that Smith is a Prime Ministerial candidate and the candidates are campaigning hard today. Smith might say:

(1) "I don't have high credence in anything that the man who will be Prime Minister says today."

Imagine further, that unknown to Smith, he himself will become Prime Minister. (1) therefore expresses a self-referential probability assertion analogous to π, the self-referential probability example from Page 680. To reject Smith's assertion of (1) as formalisable would put serious restrictions on the natural language sentences that are formalisable.

Such discussions are not new. The possibility of self-reference is also at the heart of the liar paradox, namely a sentence that says of itself that it is not true. This can be expressed by:

(λ) $\neg \mathsf{T}\ulcorner \lambda \urcorner$

This liar sentence leads to contradiction under certain basic assumptions about the truth predicate, T, namely the principles $\mathsf{T}\ulcorner \varphi \urcorner \leftrightarrow \varphi$ for each φ.

In Kripke's seminal paper, he says:

> Many, probably most, of our ordinary assertions about truth or falsity are liable, if our empirical facts are extremely unfavourable, to exhibit

[2] For some choice of values of α, for example the rational numbers.

[3] An example of empirical self-reference in the case of truth is Kripke's Nixon example from Kripke (1975, p. 695).

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

> paradoxical features.... it would be fruitless to look for an intrinsic criterion that will enable us to sieve out—as meaningless, or ill-formed—those sentences which lead to paradox. (Kripke, 1975, p. 691–692)

Analogously, if we wish our formal language to represent our ordinary assertions about probability we should allow for the possibility of self-referential sentences. We should then provide a clear syntax and semantics that can appropriately deal with these sentences as well as providing an axiomatic theory for reasoning about the language. This is what we do in this paper.

We will now briefly mention some previous work on self-referential probabilities. Although there is not space to give proper discussion of these papers we will sketch the ideas so it is clearer where our work fits in.

In Leitgeb (2012), Leitgeb develops the beginnings of what might be called a revision semantics for probability, though he only goes to stage ω. He also provides an axiomatic theory. Revision semantics for truth is a popular alternative to the semantics that we focus on. Our paper can therefore be seen to connect to and complement Leitgeb's work by seeing how the variety of theories of truth can lead to theories of probability.

In Caie (2013) and Caie (2014), Caie argues that traditional arguments for probabilism, such as the argument from accuracy, the Dutch Book argument and the argument from calibration all need to be modified in the presence of self-referential probabilities, and that so modified they do not lead to the rational requirement for beliefs to be probabilistic. Some further analysis of Caie's modifications can be found in Campbell-Moore (2015). In Caie (2013), Caie also presents a prima facie argument against probabilism by noticing its inconsistency with introspection when such self-reference is present. Our proposal in this paper, that introspection should be stated by appeal to a truth predicate, can be seen as another response to Caie's argument.

Lastly, the unpublished paper (Christiano *et al.*, n.d.) also considers the challenge that probabilism is inconsistent with introspection. In their paper, Christiano *et al.* show that probabilism is consistent with an approximate version of introspection where one can only apply introspection to open intervals of values in which the probability lies. These authors come from a computer science background and believe that these self-referential probabilities might have a role to play in the development of artificial intelligence.

Both at the final stage of Leitgeb's construction and in the construction by Christiano *et al.*, there is a formula $\varphi(x)$ such that $P(\exists x \varphi(x)) = 1$ but for each n $P(\varphi(\overline{0}) \vee \ldots \vee \varphi(\overline{n})) = 0$. This shows that they badly fail the Gaifman condition. In Section 4.2 we will show that our semantics allows P to satisfy the version of the Gaifman condition that is appropriate in our framework.

Since writing this paper we have seen that in Caie (2011, p. 64) Caie has presented a semantics that also generalises Kripke's semantics to deal with self-referential probabilities. Our construction is more general, for example we work with more general background structures and our definitions will also apply to non-consistent evaluation functions, though we do not consider these in our paper.

§3. A semantics for languages with self-referential probabilities.

3.1. Setup: Language and notataion. The syntax of the language we work with will be as follows:

DEFINITION 3.1. *Let L be some language extending the language of Peano arithmetic. We allow for the addition of contingent vocabulary but for technical ease we shall only*

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

allow contingent relation symbols (and propositional variables) and not function symbols or constants.[4] *We also only allow for a countable number of contingent vocabulary symbols in order for our language to remain countable and the completeness proof to work.*

Let $L_{\mathsf{P},\mathsf{T}}$ *extend this language by adding a unary predicate* T *and a binary predicate* $\mathsf{P}_{\geqslant}$.

We could consider languages with multiple probability notions, then we would add the binary predicate $\mathsf{P}^{\mathrm{A}}_{\geqslant}$ for each notion of probability, or agent A, but our constructions will immediately generalise to the multiple probability languages so we just focus on the language with one probability notion. We have included the truth predicate since it is easy to extend the definition of the semantics to deal with truth as well as probability and it is nice to see that the construction can give a joint theory of truth and probability. Additionally, we shall rely on the truth predicate for our later axiomatisation and for expressing principles such as introspection.

We need to be able to represent sentences of $L_{\mathsf{P},\mathsf{T}}$ as objects in the language. We therefore assume some standard Gödel coding of the $L_{\mathsf{P},\mathsf{T}}$ expressions into the natural numbers. We shall also assume a coding of rational numbers into the natural numbers, which allows us to have the technical ease of having Peano arithmetic as our background theory. We then have sentences such as $\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$ whose intended interpretation is

"The probability of φ is $\geqslant \alpha$."

Since $\mathsf{P}_{\geqslant}$ is a predicate and we are working in first order logic we can quantify over both positions of the predicate $\mathsf{P}_{\geqslant}$, so we have sentences like $\exists x \exists a \mathsf{P}_{\geqslant}(x, a)$.[5]

It is important to note that although we have a language that can only talk about rational numbers we haven't assumed that sentences always have rational probability values. For example we might have a model where for each rational $\alpha < \sqrt{2}/2$, $\mathsf{P}_{>}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$, and for each rational $\alpha > \sqrt{2}/2$, $\mathsf{P}_{<}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$. In that case the probability of φ would be said to be $\sqrt{2}/2$, which is an irrational number. The restriction is just that we cannot directly talk about this probability value because we don't have an object in our language standing for $\sqrt{2}/2$.

We now introduce the notation we will use for the construction of the semantics.[6]

NOTATION 3.2. *We assume some coding* $\# : \mathsf{Expressions}(L_{\mathsf{P},\mathsf{T}}) \cup \mathbb{Q} \to \mathbb{N}$ *that is recursive and one-to-one. For* φ *an expression of* $L_{\mathsf{P},\mathsf{T}}$, *i.e.* $\varphi \in \mathsf{Expressions}(L_{\mathsf{P},\mathsf{T}})$ *and* α *a rational number, i.e.* $\alpha \in \mathbb{Q}$, *we let* $\ulcorner\varphi\urcorner$ *and* $\ulcorner\alpha\urcorner$ *denote the numerals*[7] *corresponding to* $\#\varphi$ *and* $\#\alpha$ *respectively. We use* $\mathsf{rat}(n)$ *to denote the rational number whose code is* n. *So* $\mathsf{rat}(\#\alpha) = \alpha$. *We denote the set of codes of rational numbers by* Rat *and the set of codes of sentences of* $L_{\mathsf{P},\mathsf{T}}$ *by* $\mathsf{Sent}_{\mathsf{P},\mathsf{T}}$.[8]

---

[4] This does not place a restriction on the expressive power of the languages since one can replace constants by unary predicates and n-ary function symbols by $n+1$-ary relation symbols. This restriction could be dropped, but then $t^{\mathbb{N}}$, as will be defined in Notation 3.2, would be ill-defined and this would just complicate the presentation of the material.

[5] We shall generally use variables a and b when the intention is that they quantify over rational numbers.

[6] Some additional notation will be introduced in Notation 5.1 but since that notation will only be used in the axiomatisation we shall delay it to there to keep this section understandable.

[7] The numeral of n is denoted $\overline{n}$ and it corresponds to the expression $\overbrace{S(\ldots S}^{n}(0)\ldots)$, where S is the successor symbol from Peano arithmetic.

[8] We shall also occasionally use $\mathsf{Sent}_{\mathsf{P},\mathsf{T}}$ to denote the set of sentences of $L_{\mathsf{P},\mathsf{T}}$ (as opposed to the *codes* of the sentences), but this should not cause any confusion.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

We use $\dot{\neg}$ *to represent the syntactic operation of negating a sentence,*[9] *so* $\dot{\neg}\ulcorner\varphi\urcorner = \ulcorner\neg\varphi\urcorner$ *is a theorem of Peano arithmetic. Similarly we use* $1\dot{-}$ *to represent "*$1-$*", and* $\succ$ *to represent the ordering* $>$ *on rational numbers.*

Finally, we denote the interpretation of the term t *in the standard model of arithmetic,* $\mathbb{N}$*, by* $t^{\mathbb{N}}$*, for example* $S\overline{n}^{\mathbb{N}} = n + 1$*. This is well-defined because we assumed there were no contingent function symbols or constants in the language* L*.*

We now introduce the other probability predicates, which we use as abbreviations.

DEFINITION 3.3. *Define for terms* t *and* s *the following abbreviations:*

- $\mathsf{P}_{>}(t, s) := \exists a \succ s(\mathsf{P}_{\geqslant}(t, a))$
- $\mathsf{P}_{\leqslant}(t, s) := \mathsf{P}_{\geqslant}(\dot{\neg}t, 1\dot{-}s)$
- $\mathsf{P}_{<}(t, s) := \mathsf{P}_{>}\left(\dot{\neg}t, 1\dot{-}s\right)$
- $\mathsf{P}_{=}(t, s) := \mathsf{P}_{\geqslant}(t, s) \wedge \mathsf{P}_{\leqslant}(t, s)$

In a model that interprets the arithmetic vocabulary by the standard model of arithmetic, we will have that $\mathsf{P}_{>}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$ holds if and only if there is some rational $\beta > \alpha$ such that $\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\beta\urcorner)$ holds.

3.2. The construction of the semantics. We will now move to developing our semantics.

Kripke's construction in Kripke (1975) is motivated by the idea that one should consider the process of evaluating a sentence to determine which sentences can unproblematically be given a truth value.

To evaluate the sentence $\mathsf{T}\ulcorner 0 = 0\urcorner$ one first has to evaluate the sentence $0 = 0$. Since $0 = 0$ does not mention the concept of truth it can easily be evaluated so $\mathsf{T}\ulcorner 0 = 0\urcorner$ can then also be evaluated. Kripke formalises this process of evaluating sentences. We shall say evaluated positively (and evaluated negatively) instead of evaluated as true (and evaluated as false) to make it clear that this is happening at the meta-level.

To evaluate $\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$ we first need to evaluate φ not only in the actual state of affairs but also in some other states of affairs. We therefore base our construction on structures with multiple "possible worlds" and we evaluate the sentences at all the worlds. We will assume that each world has a "degree of accessibility" relation to the other worlds. This will be used to give us the interpretation of $\mathsf{P}_{\geqslant}$.

DEFINITION 3.4 (Probabilistic modal structure). *A* probabilistic modal structure *for a language* L *is given by a frame and an interpretation:*

A frame *is some* $(W, \{m_w | w \in W\})$ *where* W *is some non-empty set, we shall call its objects* worlds*, and* m_w *is some finitely additive probability measure over the powerset of* W*,*[10] *i.e.* $m_w : \mathcal{P}(W) \to \mathbb{R}$ *satisfying:*

- $m_w(W) = 1$
- $m_w(A) \geqslant 0$ *for all* $A \subseteq W$
- *For* $A, B \subseteq W$*, if* $A \cap B = \varnothing$ *then* $m_w(A \cup B) = m_w(A) + m_w(B)$

[9] Which is primitive recursive so is representable in Peano arithmetic.

[10] Assuming that this is defined on the whole powerset does not in fact lead to any additional restriction when we deal with merely-finitely additive probability measures, since a finitely additive probability measure on some Boolean algebra can always be extended to one defined on the whole powerset.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

An interpretation, $\mathbb{M}$, *assigns to each world* w, *a classical model for the language* L, $\mathbb{M}(w)$. *For the purpose of this paper we assume that each* $\mathbb{M}(w)$ *has the natural numbers as a domain and interprets the arithmetic vocabulary in the standard way. We call such models* $\mathbb{N}$*-models.*[11]

Such structures are very closely related to type spaces, which are of fundamental importance in game theory and economics. In a type space it is almost always assumed that each m_w is σ-additive. Furthermore it is often assumed that $m_w\{v \mid m_v = m_w\} = 1$, i.e. the agents are fully aware of their own beliefs, though these are often called Harysani type spaces following Harysani's development of them in Harsanyi (1967). We use a different name to make it clear that we do not use these assumptions. These structures can also be seen as quantitative versions of Kripke structures from modal logic.

We will now give an example of a probabilistic modal structure and will then motivate our construction. The reader familiar with such structures and the Kripke construction and who is eager to jump to our formal definition of the semantics can find that in Definition 3.6 on Page 690.

Consider the following example: Suppose we have an urn filled with 90 balls, 30 of which are yellow, 30 blue and 30 red. Suppose that a random ball is drawn from the urn and the agent is told whether it is yellow or not. We will give a probabilistic modal structure that represents the agent's degrees of belief once the ball has been drawn and she has been told whether it is yellow or not. To formalise this example we use a language, L, that adds to the language of Peano arithmetic the propositional variables Y, B and R, which will stand for the propositions that a yellow, blue or red ball is drawn, respectively. We consider three worlds that will be used to represent the colour of the ball drawn, so we take $W = \{w_\mathsf{Y}, w_\mathsf{B}, w_\mathsf{R}\}$ where w_Y is actual if a yellow ball was drawn, w_B for the blue ball and w_R for the red. The interpretation function $\mathbb{M}$ describes what these worlds are like, for example the model $\mathbb{M}(w_\mathsf{Y})$ assigns the truth-value true to Y and false to B and R. The other component we need to finish our description of the probabilistic modal structure are the functions m_w representing how much our agent thinks the other worlds are possible if she is actually in the world w. If a yellow ball is actually drawn, i.e. the agent is in the world w_Y, then she is told that the ball is yellow, so she is certain that she is in w_Y. We therefore have that $m_{w_\mathsf{Y}}(\{w_\mathsf{Y}\}) = 1$, $m_{w_\mathsf{Y}}(\{w_\mathsf{B}\}) = 0$ and $m_{w_\mathsf{Y}}(\{w_\mathsf{R}\}) = 0$. Since there are only finitely many worlds this is enough information to determine the full m_{w_Y}.[12] If a blue ball is actually drawn, i.e. the agent is in w_B, then she is told that the ball is not yellow so the only worlds she considers as still possible are w_B and w_R. The agent thinks it is as likely that a blue ball is drawn as a red ball, so we will have that $m_{w_\mathsf{B}}(\{w_\mathsf{Y}\}) = 0$, $m_{w_\mathsf{B}}(\{w_\mathsf{B}\}) = 1/2$ and $m_{w_\mathsf{B}}(\{w_\mathsf{R}\}) = 1/2$, which is again enough to determine the full m_{w_B}. The case when a red ball is actually drawn is the same from the agent's perspective as if a blue ball is actually drawn so $m_{w_\mathsf{B}} = m_{w_\mathsf{R}}$.

We can represent this probabilistic modal structure by the diagram in Fig. 1. In this example the space is finite so we can represent the measures by degree of accessibility relations, which we have done by the labelled arrows in the diagram. We have omitted the the arrows that would be labelled by 0.

---

[11] This restriction of the domain allows us to have a name for each member of the domain and therefore makes the presentation easier since we can then give the semantics without mentioning open formulas and variable assignments. This restriction also helps for the axiomatisation.

[12] Which will be given by: $m_{w_\mathsf{Y}}(A) = \sum_{w \in A} m_{w_\mathsf{Y}}(w)$.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

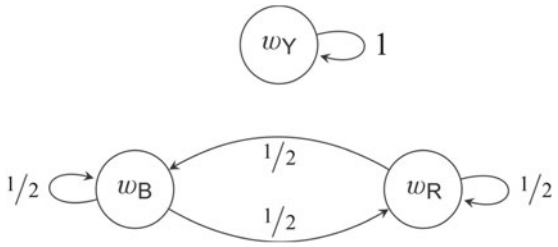


Fig. 1. Example of a probabilistic modal structure.

At the first stage we will use $\mathbb{M}$ to see that B is evaluated positively in w_{B} and negatively in the other worlds. So using the frame, we will then be able to see that at the second stage we should now evaluate $\mathsf{P}_{\geqslant}(\ulcorner\mathsf{B}\urcorner, \ulcorner 1/2\urcorner)$ positively in w_{B} and w_{R} and negatively in w_{Y}.

To formalise the evaluation procedure we need to record how the sentences have been evaluated at each world. We do this by using an *evaluation function* that records the codes of the sentences that are evaluated positively at each world. In doing this we only focus on those sentences that are evaluated positively and see that φ is evaluated negatively if and only if $\neg\varphi$ is evaluated positively.

DEFINITION 3.5. *An* evaluation function*, f, assigns to each world, w, a set $f(w) \subseteq \mathbb{N}$.*

If $\#\varphi \in f(w)$, we say that f evaluates φ positively at w.

We can now proceed to motivate the formal analysis of the evaluation procedure. We do this by developing a definition of $\Theta(f)$, which is the evaluation function given by another step of reasoning. So if f gives the codes of the sentences that we have so far evaluated positively, then $\Theta(f)$ gives the codes of the sentences that one can evaluate positively at the next stage.

At the zero-th stage one often starts without having evaluated any sentence either way. This can be given by an evaluation function f_0 with $f_0(w) = \varnothing$ for all w.

A sentence that does not involve truth or probability can be evaluated positively or negatively by just considering $\mathbb{M}(w)$. So we define:

- For φ a sentence of L, $\#\varphi \in \Theta(f)(w) \iff \mathbb{M}(w) \models \varphi$
- For φ a sentence of L, $\#\neg\varphi \in \Theta(f)(w) \iff \mathbb{M}(w) \not\models \varphi$

This will give the correct evaluations to the sentences of L, for example $0 = 0 \in \Theta(f)(w)$ and $\neg 0 = 1 \in \Theta(f)(w)$.

To evaluate a sentence $\mathsf{T}\ulcorner\varphi\urcorner$ we first evaluate φ. If φ was evaluated positively then we can now evaluate $\mathsf{T}\ulcorner\varphi\urcorner$ positively, and similarly if it was evaluated negatively. However, if φ was not evaluated either way then we still do not evaluate $\mathsf{T}\ulcorner\varphi\urcorner$ either way. This is described by the clauses:

- $\#\mathsf{T}\ulcorner\varphi\urcorner \in \Theta(f) \iff \#\varphi \in f(w)$
- $\#\neg\mathsf{T}\ulcorner\varphi\urcorner \in \Theta(f) \iff \#\neg\varphi \in f(w)$

We therefore get that $\#\mathsf{T}\ulcorner 0 = 0\urcorner \in \Theta(\Theta(f))(w)$ and $\#\neg\mathsf{T}\ulcorner 0 = 1\urcorner \in \Theta(\Theta(f))(w)$.

To describe the cases for probability we consider the fragment of a probabilistic modal frame that is pictured in Fig. 2. We consider how one should evaluate $\mathsf{P}_{\geqslant}(\ulcorner\psi\urcorner, \ulcorner\alpha\urcorner)$ for different values of α.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

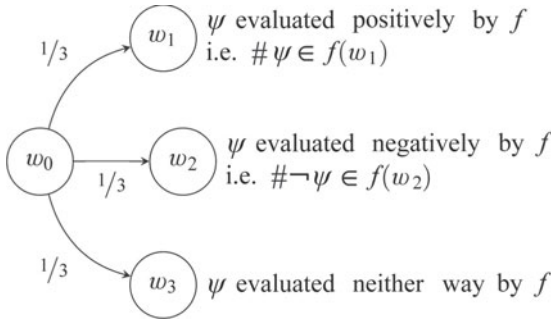


Fig. 2. A fragment of a probabilistic modal structure representing the information required to evaluate $\mathsf{P}_{\geqslant}(\ulcorner\psi\urcorner, \ulcorner\alpha\urcorner)$ in $\Theta(f)(w_0)$.

$\mathsf{P}_{\geqslant}(\ulcorner\psi\urcorner, \ulcorner 0.3\urcorner)$ will be evaluated positively by $\Theta(f)$ because the measure of the worlds where ψ is evaluated positively is $1/3 = 0.333\ldots$, which is larger than 0.3.[13] $\mathsf{P}_{\geqslant}(\ulcorner\psi\urcorner, \ulcorner 0.7\urcorner)$ will be evaluated negatively by $\Theta(f)$ because however ψ will be evaluated in w_3 there are be too many worlds where ψ is already evaluated negatively for the measure of the worlds where it is evaluated positively to become larger than 0.7. While the evaluation function remains consistent this measure could at most become $0.666\ldots = 1 - m_w\{v \mid \#\neg\psi \notin f(v)\}$. We evaluate $\mathsf{P}_{\geqslant}(\ulcorner\psi\urcorner, \ulcorner 0.5\urcorner)$ neither way because if ψ was to become evaluated in w_3 the measure of the worlds where ψ is evaluated positively would become either $0.333\ldots$ or $0.666\ldots$ so we need to retain the flexibility that $\mathsf{P}_{\geqslant}(\ulcorner\psi\urcorner, \ulcorner 0.5\urcorner)$ can later be evaluated either positively or negatively depending on how ψ is evaluated at w_3.

We therefore give the definition

- $\#\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner) \in \Theta(f)(w) \iff m_w\{v \mid \#\varphi \in f(v)\} \geqslant \alpha$
- $\#\neg\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner) \in \Theta(f)(w) \iff m_w\{v \mid \#\neg\varphi \in f(v)\} > 1 - \alpha$

In this paper we only consider consistent evaluation functions (see Definition 3.9).[14] Restricting to only consistent evaluation functions, we have that $m_w\{v \mid \#\neg\varphi \in f(v)\} > 1-\alpha$ exactly captures the requirement that the measure of the worlds where φ is evaluated positively will not become $\geqslant \alpha$ however φ becomes evaluated in the worlds where it is not currently evaluated, i.e. for all g (consistent) extending f, $m_w\{v \mid \#\varphi \in g(v)\} \not\geqslant \alpha$.

In this example we saw that the probability of ψ is given by a range. This is described pictorially in Fig. 3.

We lastly need to give the definitions for the connectives and quantifiers. For example we need to say how $\varphi \vee \neg\varphi$ should be evaluated if φ is itself evaluated neither way. For this we directly use the strong Kleene three valued evaluation scheme, which is the scheme that Kripke focused on and there has been a lot of work following him in this. Using this scheme results in having that $\#\varphi \vee \psi \in \Theta(f)(w)$ if and only if $\#\varphi \in \Theta(f)(w)$ or $\#\psi \in \Theta(f)(w)$, so if φ is evaluated neither way then $\varphi \vee \neg\varphi$ will also be evaluated neither way. The advantage of this scheme over, for example, one based on supervaluational logic is that it

[13] One should really say "the measure of *the set of* the worlds where ψ is evaluated positively", but that would be cumbersome.

[14] This definition will also apply when we want to work with non-consistent evaluation functions. This is an advantage over Caie (2011). In non-consistent evaluation functions one may have, e.g.: $\#\mathsf{P}_{\geqslant}(\ulcorner\lambda\urcorner, \ulcorner 1\urcorner) \in \Theta(f)(w)$, $\#\neg\mathsf{P}_{\geqslant}(\ulcorner\lambda\urcorner, \ulcorner 1\urcorner) \in \Theta(f)(w)$, and $\#\mathsf{P}_{\leqslant}(\ulcorner\lambda\urcorner, \ulcorner 0\urcorner) \in \Theta(f)(w)$.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

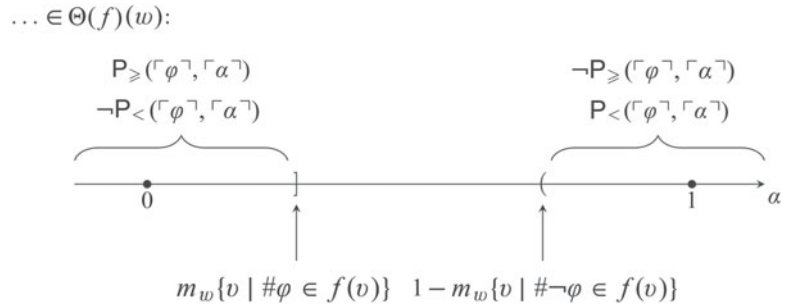


Fig. 3. How $\Theta(f)(w)$ evaluates the probability of φ.

is truth functional, so for example the evaluation of $\varphi \vee \psi$ depends only on how φ and ψ have been evaluated.

This fully defines $\Theta(f)$. We only used the question of whether φ can now be evaluated positively, i.e. if $\varphi \in \Theta(f)$, as motivating the definition. We formally understand it as a definition of a three valued semantics, $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}}$, and we will then later define $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi \iff \#\varphi \in \Theta(f)$. This is common when working with Kripke's theory of truth. We sum up our discussion in the formal definition of $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}}$.

DEFINITION 3.6. *For $\mathcal{M}$ a probabilistic modal structure, $w \in W$, and f an evaluation function, define $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}}$ by induction on the positive complexity of the formula as follows.*

- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi \iff \mathbb{M}(w) \models \varphi$ *for φ an atomic sentence of L*
- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg\varphi \iff \mathbb{M}(w) \not\models \varphi$ *for φ an atomic sentence of L*
- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{T}t \iff t^{\mathbb{N}} \in f(w)$ *and* $t^{\mathbb{N}} \in \mathsf{Sent}_{\mathsf{P},\mathsf{T}}$
- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg\mathsf{T}t \iff \dot{\neg}t^{\mathbb{N}} \in f(w)$ *or* $t^{\mathbb{N}} \notin \mathsf{Sent}_{\mathsf{P},\mathsf{T}}$
- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{\geqslant}(t, s) \iff m_w\{v \mid t^{\mathbb{N}} \in f(v)\} \geqslant \mathsf{rat}(s^{\mathbb{N}})$ *and* $s^{\mathbb{N}} \in \mathsf{Rat}$
- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg\mathsf{P}_{\geqslant}(t, s) \iff m_w\{v \mid \dot{\neg}t^{\mathbb{N}} \in f(v)\} > 1 - \mathsf{rat}(s^{\mathbb{N}})$ *or* $s^{\mathbb{N}} \notin \mathsf{Rat}$
- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg\neg\varphi \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi$
- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi \vee \psi \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi$ *or* $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \psi$
- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg(\varphi \vee \psi) \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg\varphi$ *and* $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg\psi$
- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \exists x\varphi(x) \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi[\overline{n}/x]$ *for some* $n \in \mathbb{N}$.
- $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg\exists x\varphi(x) \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg\varphi[\overline{n}/x]$ *for all* $n \in \mathbb{N}$

Note that we have omitted the connectives $\vee, \rightarrow, \leftrightarrow, \forall$. We shall take them as abbreviations, for example "$\varphi \wedge \psi$" abbreviates "$\neg(\neg\varphi \vee \neg\psi)$".

The only difference to the standard definition is the addition of the clauses for probability. As a consequence of our definition and the fact that $\mathbb{M}(w)$ is an $\mathbb{N}$-model we have that $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{>}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner) \iff m_w\{v \mid \#\varphi \in f(v)\} > \alpha$.

We now give the definition of Θ in terms of $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}}$.

DEFINITION 3.7. *Define Θ a function from evaluation functions to evaluation functions by*

$$\Theta(f)(w) := \{\#\varphi \mid (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi\}.$$

We now consider an example of how this works for the "unproblematic" sentences.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

Consider again the example in Fig. 1, which models an agent's beliefs after a ball is picked from an urn and the agent is told whether it's yellow or not.

Take any f. Observe that:

$$(w_{\mathsf{B}}, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{B} \text{ and } (w_{\mathsf{R}}, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg\mathsf{B}$$

so:

$$\#\mathsf{B} \in \Theta(f)(w_{\mathsf{B}}) \text{ and } \#\neg\mathsf{B} \in \Theta(f)(w_{\mathsf{R}}).$$

Therefore:

$$(w_{\mathsf{B}}, \Theta(f)) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{=}(\ulcorner\mathsf{B}\urcorner, \ulcorner 1/2\urcorner) \text{ and similarly for } w_{\mathsf{R}}.^{15}$$

so:

$$\#\mathsf{P}_{=}(\ulcorner\mathsf{B}\urcorner, \ulcorner 1/2\urcorner) \in \Theta(\Theta(f))(w_{\mathsf{B}}) \text{ and similarly for } w_{\mathsf{R}}.$$

Then by similar reasoning:

$$(w_{\mathsf{B}}, \Theta(\Theta(f))) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{=}(\ulcorner\mathsf{P}_{=}(\ulcorner\mathsf{B}\urcorner, \ulcorner 1/2\urcorner)\urcorner, \ulcorner 1\urcorner)$$

so:

$$\#\mathsf{P}_{=}(\ulcorner\mathsf{P}_{=}(\ulcorner\mathsf{B}\urcorner, \ulcorner 1/2\urcorner)\urcorner, \ulcorner 1\urcorner) \in \Theta(\Theta(\Theta(f)))(w_{\mathsf{B}}).$$

These sentences have an easy translation into the operator language. Such sentences will be given point-valued probabilities and be evaluated positively or negatively by some $\Theta(\Theta(\ldots\Theta(f)\ldots))$. This shows that this semantics extends an operator semantics, a minimal adequacy requirement for any proposed semantics.[16]

If one starts with $f(w) = \varnothing$ for each $w \in W$, and iteratively applies Θ, then Θ will only give evaluations to sentences that were previously evaluated neither way, it will not *change* the evaluation of a sentence. This is because Θ is monotone:

LEMMA 3.8 (Θ is monotone). *If for all w $f(w) \subseteq g(w)$, then also for all w $\Theta(f)(w) \subseteq \Theta(g)(w)$.*

Proof. Take some evaluation functions f and g such that $f(w) \subseteq g(w)$ for all w. It suffices to prove that if $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi$ then $(w, g) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi$. We do this by induction on the positive complexity of φ. □

This fact ensures that there are fixed points of the operator Θ, i.e., evaluation functions f with $f = \Theta(f)$. These are evaluation functions where the process of evaluation doesn't lead to any new "information".

DEFINITION 3.9. *f is called a* fixed point *evaluation function if $\Theta(f) = f$.*

f is called consistent *if for each $w \in W$ and $n \in \mathbb{N}$, it is not the case that $n \in f(w)$ and $\neg\dot{n} \in f(w)$.*

COROLLARY 3.10 (Θ has fixed points). *For every $\mathcal{M}$ there is some consistent fixed point evaluation function f.*[17]

---

[15] Remember "$\mathsf{P}_{=}(\ulcorner\mathsf{B}\urcorner, \ulcorner 1/2\urcorner)$" is an abbreviation for "$\mathsf{P}_{\geqslant}(\ulcorner\mathsf{B}\urcorner, \ulcorner 1/2\urcorner) \wedge \mathsf{P}_{\geqslant}(\ulcorner\neg\mathsf{B}\urcorner, \ulcorner 1 - 1/2\urcorner)$".

[16] We can show for the natural translation function ρ from the operator language as presented in Heifetz & Mongi (2001) to the predicate language, and f a fixed point one has

$$w \models_{\mathcal{M}} \varphi \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \rho(\varphi).$$

[17] To show that the minimal fixed point is consistent one works by induction on the generation procedure by showing that if f is consistent then so is $\Theta(f)$.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

If φ is grounded in facts that are not about truth or probability then this process of evaluation will terminate in such facts and the sentence will be evaluated appropriately in a fixed point of Θ. Such sentences will also therefore be given a point-valued probability, as is desired. For example, $0 = 0 \vee \lambda$ will be evaluated positively in each world and so be assigned probability 1, i.e. $\mathsf{P}_{=}(\ulcorner 0 = 0 \vee \lambda \urcorner, \ulcorner 1 \urcorner)$ will also be evaluated positively, even though $0 = 0 \vee \lambda$ is not expressible in the operator language.

The fixed points have some nice properties.

PROPOSITION 3.11. *For f a fixed point of Θ we have:*

$$\#\varphi \in f(w) \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi$$

Therefore we have

$$(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{T}\ulcorner\varphi\urcorner \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi$$

$$(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner) \iff m_w\{v \mid (v, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi\} \geqslant \alpha$$

Since $\mathbb{M}(w)$ is an $\mathbb{N}$-model we also have:

$$(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{>}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner) \iff m_w\{v \mid (v, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi\} > \alpha$$

Proof. Follows immediately from Definitions 3.6 and 3.7. □

3.3. The classical semantics. We do not propose "$(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}}$" as the semantics for the language, instead we use the interpretation of T and P that $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}}$ gives us to determine a classical model for the language $L_{\mathsf{P},\mathsf{T}}$. This is common when working with Kripke's theory.

We will define *the induced model given by $\mathcal{M}$ and f at w*, $\mathsf{IndMod}_{\mathcal{M}}[w, f]$, by "closing off" the model by putting the unevaluated sentences outside of the extension of T and $\mathsf{P}_{\geqslant}$. This is described pictorially by altering Fig. 3 to Fig. 4.

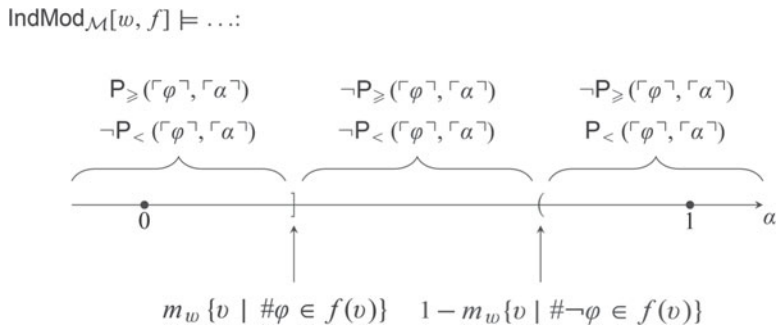


Fig. 4. How $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ evaluates the probability of φ.

It is defined formally as follows:

DEFINITION 3.12. *Define $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ to be a (classical) model for the language $L_{\mathsf{P},\mathsf{T}}$ that has the domain $\mathbb{N}$, interprets the predicates from L as is specified by $\mathbb{M}(w)$, and interprets the other predicates by:*

- $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{T}\overline{n} \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{T}\overline{n}$
- $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{P}_{\geqslant}(\overline{n}, \overline{m}) \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{\geqslant}(\overline{n}, \overline{m})$

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

This will satisfy:

PROPOSITION 3.13. *For $\mathcal{M}$ a probabilistic modal structure, f an evaluation function and $w \in W$,*

- $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \varphi \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi$, *for φ a sentence of L*
- $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{P}_{>}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner) \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{>}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$
- $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{P}_{\leqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner) \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{\leqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$
- $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{P}_{<}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner) \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{<}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$
- $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{P}_{=}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner) \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{=}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$

Although these equivalences hold, $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models$ differs from $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}}$ because $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ is classical, for example we might have that $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \neg\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$ but $(w, f) \not\models^{\mathsf{SKP}}_{\mathcal{M}} \neg\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$. These induced models are classical models, but the "inner logic of T"[18] will not be classical and not every sentence will be assigned a point valued probability. For example for f a consistent fixed point we will have that $\neg\mathsf{T}\ulcorner\lambda\urcorner$, $\neg\mathsf{T}\ulcorner\neg\lambda\urcorner$, $\neg\mathsf{P}_{>}(\ulcorner\lambda\urcorner, \ulcorner 0\urcorner)$ and $\neg\mathsf{P}_{<}(\ulcorner\lambda\urcorner, \ulcorner 1\urcorner)$ all satisfied in $\mathsf{IndMod}_{\mathcal{M}}[w, f]$.

These induced models, for consistent fixed points f, are our proposal for the semantics of the language.

§4. Observations and comments on the semantics.

4.1. Introspection. Studying introspection in languages that allow for self-referential probabilities is interesting because if it is naively formulated it is inconsistent, a problem discussed in Caie (2013) and Christiano *et al.* (n.d.).

A probabilistic modal structure that has the property that:

$$\text{For all } w, m_w\{v | m_v = m_w\} = 1$$

will satisfy introspection in the operator language. That is:

$$\mathrm{P}_{\geqslant \alpha}(\varphi) \rightarrow \mathrm{P}_{=1}(\mathrm{P}_{\geqslant \alpha}(\varphi)),$$

$$\neg\mathrm{P}_{\geqslant \alpha}(\varphi) \rightarrow \mathrm{P}_{=1}(\neg\mathrm{P}_{\geqslant \alpha}(\varphi)).$$

Such probabilistic modal structures will also satisfy introspection in the predicate setting if the principles are expressed using a truth predicate.

PROPOSITION 4.1. *Let $\mathcal{M}$ be such that $m_w\{v \mid m_v = m_w\} = 1$ for all w. Then for any evaluation function f and world w,*

- *If* $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$ *then* $(w, \Theta(f)) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{=}(\ulcorner\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)\urcorner, \ulcorner 1\urcorner)$
- *If* $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \neg\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)$ *then* $(w, \Theta(f)) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{=}(\ulcorner\neg\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner)\urcorner, \ulcorner 1\urcorner)$

And similarly for $\mathsf{P}_{>}$, $\mathsf{P}_{<}$, $\mathsf{P}_{\leqslant}$ *and* $\mathsf{P}_{=}$.

[18] By this we mean the logical laws that hold in the inside applications of the truth predicate.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

By the definition of $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ *we therefore have:*[19]

- $\mathsf{IndMod}_{\mathcal{M}}[w, \Theta(f)] \models \mathsf{T}^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner} \rightarrow \mathsf{P}_{=}(^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})\big)$
- $\mathsf{IndMod}_{\mathcal{M}}[w, \Theta(f)] \models \mathsf{T}^{\ulcorner}\neg\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner} \rightarrow \mathsf{P}_{=}(^{\ulcorner}\neg\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})\big)$

And similarly for $\mathsf{P}_{>}$ *etc.*

Therefore for f *a fixed point evaluation function,*

- $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{T}^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner} \rightarrow \mathsf{P}_{=}(^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})\big)$
- $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{T}^{\ulcorner}\neg\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner} \rightarrow \mathsf{P}_{=}(^{\ulcorner}\neg\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})\big)$

And similarly for $\mathsf{P}_{>}$ *etc.*[20]

One might give an explanation for this as follows. To answer the question of whether

$$\mathsf{IndMod}_{\mathcal{M}}[w, \Theta(f)] \models \mathsf{T}^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner} \rightarrow \mathsf{P}_{\geqslant}(^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})\,?$$

one needs only to answer the question "$\#\varphi \in f(v)$?" and use the definitions. However, to answer the question of whether

$$\mathsf{IndMod}_{\mathcal{M}}[w, \Theta(f)] \models \mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner}) \rightarrow \mathsf{P}_{\geqslant}(^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})\,?$$

one needs to answer both the questions "$\#\varphi \in \Theta(f)(v)$?" and "$\#\varphi \in f(v)$?" and use the definitions.

This shows that when one asks for the version with the truth predicate, one only asks about properties of the probabilistic modal structure and not about how Θ works. This is therefore a well-motivated way to express the principle when such semantics are employed.

This is an example of a more general strategy one might employ in formulating desiderata such as introspection. The strategy comes from Stern (2014a) where he suggests that following the strategy of "avoiding introduction and elimination of the modal predicate independently of the truth predicate" might allow one to avoid paradoxes. Moreover he says:

> "[This strategy] seems to be well motivated if one adopts the deflationist idea that quotation and disquotation are the function of the truth predicate. Consequently, quotation and disquotation of sentences is not the task of the modal predicate and in formulating modal principles we should therefore avoid the introduction or elimination of the modal predicates without the detour via the truth predicate."

So if one accepts this as an appropriate formulation of introspection we have that introspection and probabilism are compatible.[21]

---

[19] In fact the quantified versions

- $\mathsf{IndMod}_{\mathcal{M}}[w, \Theta(\Theta(f))] \models \forall a \forall x(\mathsf{TP}_{\dot{\geqslant}}(x, a) \rightarrow \mathsf{P}_{\geqslant}(\mathsf{P}_{\dot{\geqslant}}(x, a), ^{\ulcorner}1^{\urcorner}))$
- $\mathsf{IndMod}_{\mathcal{M}}[w, \Theta(\Theta(f))] \models \forall a \forall x(\mathsf{T}\dot{\neg}\mathsf{P}_{\dot{\geqslant}}(x, a) \rightarrow \mathsf{P}_{\geqslant}(\dot{\neg}\mathsf{P}_{\dot{\geqslant}}(x, a), ^{\ulcorner}1^{\urcorner}))$

are satisfied but we do not present this because it is not important for our point and we don't feel that it is worth yet introducing this general $\dot{\rhd}$ notation, which is introduced in Notation 5.1.

[20] We could equivalently formalise the principles as

- $\mathsf{P}_{\geqslant}(^{\ulcorner}\mathsf{T}^{\ulcorner}\varphi^{\urcorner\urcorner}, ^{\ulcorner}\alpha^{\urcorner}) \rightarrow \mathsf{P}_{=}(^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner}))$
- $\neg\mathsf{P}_{\geqslant}(^{\ulcorner}\mathsf{T}^{\ulcorner}\varphi^{\urcorner\urcorner}, ^{\ulcorner}\alpha^{\urcorner}) \rightarrow \mathsf{P}_{=}(^{\ulcorner}\neg\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner}))$.

[21] In fact, in such probabilistic modal structures, at fixed points f, the "positive" principles $\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner}) \rightarrow \mathsf{P}_{=}(^{\ulcorner}\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})$ and $\mathsf{P}_{<}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner}) \rightarrow \mathsf{P}_{=}(^{\ulcorner}\mathsf{P}_{<}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})$ are satisfied. However, for consistent fixed points, the "negative"

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

Further work should be done to see how one should state other principles and whether doing so allows one to avoid paradoxical contradictions arising from the self-referential nature of the language.

We next show a nice feature of the construction, namely that it can account for the Gaifman condition.

4.2. Gaifman condition. A function $F : \mathsf{Sent} \to \mathbb{R}$ is said to satisfy the *Gaifman condition* if

$$\text{For all } \varphi,\ F(\exists x \varphi(x)) = \lim_{n\to\infty} F(\varphi(\overline{0}) \vee \varphi(\overline{1}) \vee \ldots \vee \varphi(\overline{n}))$$

This in part captures the idea that the domain is exactly 0, 1, 2 . . .. It was called σ-additivity in Leitgeb (2008) and Leitgeb (2012).

As was mentioned in the introduction, both Leitgeb (2012) and Christiano *et al.* (n.d.) face a challenge from the Gaifman condition because both Christiano's requirements and the final stage of Leitgeb's construction lead to a formula $\varphi(x)$ such that for each n $\mathsf{P}^{\ulcorner}\varphi(\overline{0}) \vee \ldots \vee \varphi(\overline{n})^{\urcorner} = 0$ but $\mathsf{P}^{\ulcorner}\exists x \varphi(x)^{\urcorner} = 1$.[22]

Our theory does not have this flaw. However, since our sentences are sometimes given ranges of probabilities instead of points, we should reformulate the definition of the Gaifman condition to apply to ranges. Let $\underline{\mathbf{P}}$ and $\overline{\mathbf{P}}$ denote the upper and lower bounds of the range of probabilities assigned to φ. More carefully:

DEFINITION 4.2. *Fix some probabilistic modal structure $\mathcal{M}$, evaluation function f and world w. Define*

$$\underline{\mathbf{P}}(\varphi) := \sup\{\alpha \mid \mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})\}$$

$$\overline{\mathbf{P}}(\varphi) := \inf\{\alpha \mid \mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{P}_{<}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})\}$$

This can be seen as in Fig. 5.

Using this definition (also by comparing Fig. 5 to Fig. 4) we have that $\underline{\mathbf{P}}(\varphi) = m_w\{v \mid \#\varphi \in f(v)\}$ and $\overline{\mathbf{P}}(\varphi) = 1 - m_w\{v \mid \#\neg\varphi \in f(v)\}$.

DEFINITION 4.3. *We say that* P *as given by* $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ *satisfies* the extended Gaifman condition *if* $\underline{\mathbf{P}}(\exists x \varphi(x)) = \lim_{n\to\infty} \underline{\mathbf{P}}(\varphi(\overline{0}) \vee \ldots \vee \varphi(\overline{n}))$ *and similarly for* $\overline{\mathbf{P}}$.

If we consider a probabilistic modal structure where the measure m_w is σ-additive then the extended Gaifman condition will be satisfied.

---

versions, $\neg\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner}) \to \mathsf{P}_{=}(^{\ulcorner}\neg\mathsf{P}_{\geqslant}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})$ and $\neg\mathsf{P}_{<}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner}) \to \mathsf{P}_{=}(^{\ulcorner}\neg\mathsf{P}_{<}(^{\ulcorner}\varphi^{\urcorner}, ^{\ulcorner}\alpha^{\urcorner})^{\urcorner}, ^{\ulcorner}1^{\urcorner})$ cannot be adopted for the problematic instances. We therefore see that restricting to only positive assertions of introspection is an alternative way of showing that introspection can remain consistent in this setting. We don't suggest this strategy because it is not a systematic solution.

[22] The failure of the Gaifman condition in Leitgeb's theory is closely related to McGee's ω-inconsistency result from McGee (1985). For Leitgeb a sentence displaying the bad failure of the Gaifman condition is: $\neg\overbrace{\mathsf{T}^{\ulcorner}\mathsf{T}^{\ulcorner}\ldots\mathsf{T}^{\ulcorner}\delta^{\urcorner}\ldots^{\urcorner\urcorner}}^{n+1}$ where δ is the McGee sentence, namely is a sentence with the property that $\mathsf{PA}^{L_{\mathsf{P},\mathsf{T}}} \vdash \delta \leftrightarrow \exists n\neg\overbrace{\mathsf{T}^{\ulcorner}\mathsf{T}^{\ulcorner}\ldots\mathsf{T}^{\ulcorner}\delta^{\urcorner}\ldots^{\urcorner\urcorner}}^{n+1}$. For Christiano *et al.* this is given by $\mathsf{P}^{\ulcorner}\epsilon^{\urcorner} \leqslant 1 - 1/n$ where ϵ is a sentence with the property $\mathsf{PA}^{L_{\mathsf{P},\mathsf{T}}} \vdash \epsilon \leftrightarrow \mathsf{P}^{\ulcorner}\epsilon^{\urcorner} < 1$; the fact the Christiano *et al.* face a challenge from the Gaifman condition was pointed out to me by Hannes Leitgeb.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

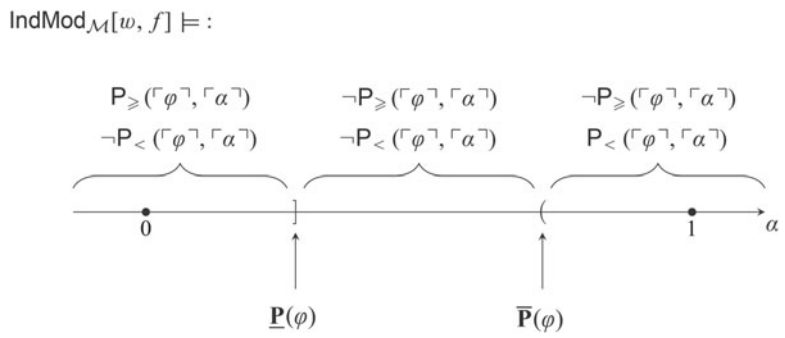


Fig. 5. Definition of $\underline{\mathbf{P}}(\varphi)$ and $\overline{\mathbf{P}}(\varphi)$.

THEOREM 4.4. *If $\mathcal{M}$ is such that m_w is σ-additive,*[23] *and f is a fixed point, then* P *as given by* $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ *will satisfy the extended Gaifman condition.*

This is the form of the Gaifman condition that is appropriate in a context where one deals with interval-valued probabilities. We therefore see that if we don't restrict ourselves to merely finitely-additive probabilities then we can account for the Gaifman condition.

4.3. P ***is an*** SK***-probability.*** Although the models $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ are classical they loose aspects of traditional probability functions. $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ does not assign particular values for the probability of φ but can be seen instead as providing ranges. As such $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ provides us with two functions to consider, these are $\underline{\mathbf{P}}$ and $\overline{\mathbf{P}}$ as given in Definition 4.2. Both these functions loose nice properties one would expect from classical probabilities, for example $\underline{\mathbf{P}}(\lambda \vee \neg\lambda) = 0$, and $\overline{\mathbf{P}}(\lambda \wedge \neg\lambda) = 1$. However we can show that $\underline{\mathbf{P}}$ and $\overline{\mathbf{P}}$ are non-classical probabilities in the sense of Williams (2014) over logics that arise from Kleene's strong three-valued evaluation scheme. In particular, $\underline{\mathbf{P}}$ is a non-classical probability over Kleene logic K_3, which is defined by truth preservation in Kleene evaluations, and $\overline{\mathbf{P}}$ is a non-classical probability over LP-logic, which is defined by falsity anti-preservation in Kleene evaluations.

§5. An axiomatic system. In the last section of this paper we present an axiomatic theory for this semantic construction. This will allow one to better reason about this semantics.

To present this we need to provide some more notation that is added to Notation 3.2.

NOTATION 5.1. *We represent the interpretation function* $^{\mathbb{N}}$ *function by* $^{\circ}$*, but this is understood not to be a function symbol in our language. We therefore have that for any closed term t, $\ulcorner t^{\mathbb{N}\circ} = \overline{\ulcorner t^{\mathbb{N}}\urcorner}$ is a non-atomic theorem of Peano arithmetic.*

We shall use Rat*,* $\mathsf{Sent}_{\mathsf{PT}}$*, and* Cterm *to represent, in our object language, the set of codes of rational numbers, sentences of L_{PT}, and closed terms of L_{PT} respectively. If $\triangleright$ is a syntactic operation we shall assume we have a function symbol $\dot{\triangleright}$ in our language representing it.*[24] *For example $\ulcorner\varphi\urcorner \dot{\vee} \ulcorner\psi\urcorner = \ulcorner\varphi \vee \psi\urcorner$ is a theorem of Peano arithmetic. Exceptions are the substitution function, which we represent by $x(y/v)$, and $^{\circ}$, which we*

[23] We can drop the condition that m_w be defined on the whole powerset of W and instead ask just that it is defined an algebra of subsets containing the sets of the form $\{v \mid n \in f(v)\}$.

[24] Observe that these are not contingent function symbols so this ensures that $t^{\mathbb{N}}$ is still well-defined.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

already introduced. We shall similarly represent operations on the rationals, for example we have a function symbol $\dot{+}$.

We now present the axiomatic system.

DEFINITION 5.2. *Remember we introduced the following abbreviations:*

- $\mathsf{P}_{>}(t,s) := \exists a \succ s(\mathsf{P}_{\geqslant}(t,a))$
- $\mathsf{P}_{\leqslant}(t,s) := \mathsf{P}_{\geqslant}(\dot{\neg}t, 1\dot{-}s)$
- $\mathsf{P}_{<}(t,s) := \mathsf{P}_{>}\left(\dot{\neg}t, 1\dot{-}s\right)$
- $\mathsf{P}_{=}(t,s) := \mathsf{P}_{\geqslant}(t,s) \wedge \mathsf{P}_{\leqslant}(t,s)$

Define ProbKFC *to be given by the following axioms, added to an axiomatisation of classical logic.*

- KFC*, the axioms for truth: (The "*C*" stands for the addition of the consistency axiom, 13.)*
  1. $\mathsf{PA}^{L_{\mathsf{P},\mathsf{T}}}$*, the axioms of Peano arithmetic with the induction schema extended to* $L_{\mathsf{P},\mathsf{T}}$.
  2. $\forall x \forall y((\mathsf{Cterm}(x) \wedge \mathsf{Cterm}(y)) \rightarrow (\mathsf{T}x \doteq y \leftrightarrow x^{\circ} = y^{\circ}))$
  3. $\forall x \forall y((\mathsf{Cterm}(x) \wedge \mathsf{Cterm}(y)) \rightarrow (\mathsf{T}\dot{\neg}x \doteq y \leftrightarrow \neg x^{\circ} = y^{\circ}))$
  4. $\forall x_1 \ldots \forall x_n((\mathsf{Cterm}(x_1) \wedge \ldots \wedge \mathsf{Cterm}(x_n)) \rightarrow (\mathsf{T}\dot{Q}x_1 \ldots x_n \leftrightarrow Qx_1^{\circ} \ldots x_n^{\circ}))$ *for each n-ary predicate* Q *of* L
  5. $\forall x_1 \ldots \forall x_n((\mathsf{Cterm}(x_1) \wedge \ldots \wedge \mathsf{Cterm}(x_n)) \rightarrow (\mathsf{T}\dot{\neg}\dot{Q}x_1 \ldots x_n \leftrightarrow \neg Qx_1^{\circ} \ldots x_n^{\circ}))$ *for each n-ary predicate* Q *of* L
  6. $\forall x(\mathsf{Sent}_{\mathsf{P},\mathsf{T}}(x) \rightarrow (\mathsf{T}\dot{\neg}\dot{\neg}x \leftrightarrow \mathsf{T}x)$
  7. $\forall x \forall y(\mathsf{Sent}_{\mathsf{P},\mathsf{T}}(x \dot{\vee} y) \rightarrow (\mathsf{T}x \dot{\vee} y \leftrightarrow (\mathsf{T}x \vee \mathsf{T}y))$
  8. $\forall x \forall y(\mathsf{Sent}_{\mathsf{P},\mathsf{T}}(x \dot{\vee} y) \rightarrow (\mathsf{T}\dot{\neg}x \dot{\vee} y \leftrightarrow (\mathsf{T}\dot{\neg}x \wedge \mathsf{T}\dot{\neg}y))$
  9. $\forall x(\mathsf{Sent}_{\mathsf{P},\mathsf{T}}(\dot{\exists}vx) \rightarrow (\mathsf{T}\dot{\exists}vx \leftrightarrow \exists y\mathsf{T}(x(y/v)))$
  10. $\forall x(\mathsf{Sent}_{\mathsf{P},\mathsf{T}}(\dot{\exists}vx) \rightarrow (\mathsf{T}\dot{\neg}\dot{\exists}vx \leftrightarrow \forall y\mathsf{T}(\dot{\neg}x(y/v)))$
  11. $\forall x(\mathsf{Cterm}(x) \rightarrow (\mathsf{T}\dot{\mathsf{T}}x \leftrightarrow \mathsf{T}x^{\circ}))$
  12. $\forall x(\mathsf{Cterm}(x) \rightarrow (\mathsf{T}\dot{\neg}\dot{\mathsf{T}}x \leftrightarrow (\mathsf{T}\dot{\neg}x^{\circ} \vee \neg\mathsf{Sent}_{\mathsf{P},\mathsf{T}}(x^{\circ})))$
  13. $\forall x(\mathsf{Sent}_{\mathsf{P},\mathsf{T}}(x) \rightarrow \neg(\mathsf{T}x \wedge \mathsf{T}\dot{\neg}x))$[25]
- InteractionAx*, the axioms for the interaction of truth and probability:*[26]
  14. $\forall x \forall y((\mathsf{Cterm}(x) \wedge \mathsf{Cterm}(y)) \rightarrow (\mathsf{T}\dot{\mathsf{P}}_{\geqslant}(x,y) \leftrightarrow \mathsf{P}_{\geqslant}(x^{\circ}, y^{\circ})))$
  15. $\forall x \forall y((\mathsf{Cterm}(x) \wedge \mathsf{Cterm}(y)) \rightarrow (\mathsf{T}\dot{\neg}\dot{\mathsf{P}}_{\geqslant}(x,y) \leftrightarrow (\mathsf{P}_{<}(x^{\circ}, y^{\circ}) \vee \neg\mathsf{Rat}(y^{\circ}))))$
- *The axioms that give basic facts about* $\mathsf{P}_{\geqslant}$*:*
  16. $\forall a(\exists x\mathsf{P}_{\geqslant}(x,a) \rightarrow \mathsf{Rat}(a))$
  17. $\forall x(\mathsf{P}_{>}(x, \ulcorner 0 \urcorner) \rightarrow \mathsf{Sent}_{\mathsf{P},\mathsf{T}}x)$
  18. $\forall x \forall a(\mathsf{Rat}(a) \rightarrow (\mathsf{P}_{\geqslant}(x,a) \leftrightarrow \forall b \prec a\mathsf{P}_{\geqslant}(x,b)))$
- *Axioms and a rule that say that* P *acts like a probability:*[27]
  19. $\mathsf{P}_{\geqslant}(\ulcorner 0 = 0 \urcorner, \ulcorner 1 \urcorner) \wedge \neg\mathsf{P}_{>}(\ulcorner 0 = 0 \urcorner, \ulcorner 1 \urcorner)$

[25] If one wished to drop this and also consider inconsistent fixed points one should replace this axiom with $\forall x(\mathsf{T}x \rightarrow \mathsf{Sent}_{\mathsf{P},\mathsf{T}}(x))$.

[26] These should be seen as the appropriate way of extending KFC to the language $L_{\mathsf{P},\mathsf{T}}$, but we include them separately to highlight them.

[27] We use the axioms for 2-additive Choquet capacities because our underlying structure might be a lattice not a boolean algebra.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

$$20\ \ \mathsf{P}_{\geqslant}(\ulcorner\neg 0 = 0\urcorner, \ulcorner 0\urcorner) \wedge \neg\mathsf{P}_{>}(\ulcorner\neg 0 = 0\urcorner, \ulcorner 0\urcorner)$$

$$21\ \ \forall x \forall y(\mathsf{Sent}_{\mathsf{P},\mathsf{T}}(x) \wedge \mathsf{Sent}_{\mathsf{P},\mathsf{T}}(y) \rightarrow \forall a\left(\mathsf{Rat}(a) \rightarrow \left(\begin{array}{l}(\forall b \forall c(\ \mathsf{P}_{\geqslant}(x,b)\ \wedge\ \mathsf{P}_{\geqslant}(y,c)\ \rightarrow b \dot{+} c \preceq a)) \\ \leftrightarrow (\forall d \forall e(\mathsf{P}_{\geqslant}(x \dot{\wedge} y, d) \wedge \mathsf{P}_{\geqslant}(x \dot{\vee} y, e) \rightarrow d \dot{+} e \preceq a))\end{array}\right)\right)$$

$$22\ \ \frac{\mathsf{T}t \rightarrow \mathsf{T}s}{\forall a(\mathsf{P}_{\geqslant}(t,a) \rightarrow \mathsf{P}_{\geqslant}(s,a))}$$

We say $\Gamma \vdash_{\mathsf{ProbKFC}} \varphi$ *if rule 22 is used* before *any members of* Γ *are used in the proof.*[28]

These axioms are sound, i.e. all induced models satisfy the axiomatisation.

THEOREM 5.3 (Soundness of ProbKFC). *Let* $\mathcal{M}$ *be a probabilistic structure,* f *a consistent fixed point and* $w \in W$*, and suppose* $\Gamma \vdash_{\mathsf{ProbKFC}} \varphi$*, then*

$$\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \Gamma \implies \mathsf{IndMod}_{\mathcal{M}}[w, f] \models \varphi.$$

Proof. By induction on the length of the proof in ProbKFC. Many of the axioms follow from Definition 3.6 using the fact that since f is a fixed point

$$\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{T}\ulcorner\varphi\urcorner \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi. \qquad \square$$

We would additionally like to have a completeness component to the axiomatisation. To get a completeness theorem we will add an ω-rule to the axiomatisation. This allows one to conclude $\forall x \varphi(x)$ from all the instances of $\varphi(\overline{n})$. It is needed to fix the standard model of arithmetic, which we used when building the semantics.

DEFINITION 5.4. *Let* ProbKFC$^{\omega}$ *be the system* ProbKFC *with the* ω*-rule added. This is a rule:*[29]

$$\frac{\varphi(\overline{0}) \qquad \varphi(\overline{1}) \qquad \varphi(\overline{2}) \qquad \varphi(\overline{3}) \qquad \ldots}{\forall x \varphi(x)}$$

We say $\Gamma \vdash^{\omega}_{\mathsf{ProbKFC}} \varphi$ *if rule 22 is used before any members of* Γ*.*

THEOREM 5.5 (Soundness and completeness). $\Gamma \vdash^{\omega}_{\mathsf{ProbKFC}} \varphi$ *if and only if for each probabilistic modal structure* $\mathcal{M}$*, consistent fixed point* f*, and* $w \in W$*,*

$$\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \Gamma \implies \mathsf{IndMod}_{\mathcal{M}}[w, f] \models \varphi$$

One could also include other "global" axioms that could be used before rule 22. These would capture facts about specific kinds of probabilistic modal structures and fixed points. The proof of the more general result allowing global axioms Σ is directly analogous but modifying the defined canonical model of Definition 5.8 by replacing $\vdash^{\omega}_{\mathsf{ProbKFC}}$ by $\vdash^{\omega}_{\mathsf{ProbKFC}\cup\Sigma}$.

[28] Rule 22 is treated like the rule of necessitation in modal logic. We have this restriction because we need $\mathsf{T}t \rightarrow \mathsf{T}s$ to hold in *all* worlds in order to deduce that $\forall a(\mathsf{P}_{\geqslant}(t,a) \rightarrow \mathsf{P}_{\geqslant}(s,a))$ is satisfied in world w. Γ includes "local assumptions" that may only hold in world w so want $\mathsf{T}t \rightarrow \mathsf{T}s$ to be derived before such local assumptions are used in order to ensure that it holds at all $v \in W$.

[29] The ω-rule can be used anywhere in the proof as opposed to rule 22 which can only be used before any assumptions (see Footnote 28). We cannot state the ω-rule as an axiom simply because we don't have the syntactic resources to do so because our language is finitary.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

We also have other forms of soundness and completeness results:

COROLLARY 5.6. *The following are equivalent:*

- $\mathcal{A} \models \Gamma \implies \mathcal{A} \models \varphi$, *whenever* $\Gamma \vdash^{\omega}_{\mathsf{ProbKFC}} \varphi$,
- *There is a probabilistic structure* $\mathcal{M}$, *consistent fixed point* f *and* $w \in W$ *such that* $\mathcal{A}$ *is elementarily equivalent to* $\mathsf{IndMod}_{\mathcal{M}}[w, f]$.[30]

Suppose $\mathcal{A}$ *is an* $\mathbb{N}$*-model.*[31] *Then the following are equivalent:*

- $\mathcal{A} \models \varphi$ *whenever* $\vdash^{\omega}_{\mathsf{ProbKFC}} \varphi$,
- *There is a probabilistic structure* $\mathcal{M}$, *consistent fixed point* f, *and* $w \in W$ *such that* $\mathcal{A} = \mathsf{IndMod}_{\mathcal{M}}[w, f]$.

The second of these results shows we have developed what Fischer *et al.* (2015) call an $\mathbb{N}$-categorical axiomatisation, although they would not consider the ω-rule as a permissible axiom as they only consider recursively enumerable theories. Theorem 5.13 could also be taken to show that $\mathsf{KFC} \cup \mathsf{InteractionAx}$, which is a recursively enumerable theory, is an $\mathbb{N}$-categorical axiomatisation *given the structure* $\mathcal{M}$.

This completeness result is proved by a canonical model construction. The fact that we can produce a canonical model is independently interesting since it gives a systematic structure one can use when working with these semantics.

5.1. Proof of the soundness and completeness of $\mathsf{ProbKFC}^{\omega}$**.** We quickly mention the soundness result of $\mathsf{ProbKFC}^{\omega}$, before moving on to sketch a proof of the completeness component.

THEOREM 5.7 (Soundness). *Let* $\mathcal{M}$ *be a probabilistic structure,* f *a consistent fixed point and* $w \in W$, *and suppose* $\Gamma \vdash^{\omega}_{\mathsf{ProbKFC}} \varphi$, *then*

$$\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \Gamma \implies \mathsf{IndMod}_{\mathcal{M}}[w, f] \models \varphi.$$

Proof. Generalise the argument in Theorem 5.3 by transfinite induction on the proof procedure. □

We can now turn to the more interesting completeness direction.

DEFINITION 5.8. *Define a probabilistic structure* $\mathcal{M}_{\mathsf{c}}$ *and evaluation function* f_{c} *as follows:*

- $W_{\mathsf{c}} := \left\{ w \subseteq \mathsf{Sent}_{\mathsf{P,T}} \;\middle|\; \begin{array}{l} w \text{ is maximally finitely } \vdash^{\omega}_{\mathsf{ProbKFC}}\text{-consistent,}^{32} \\ w \text{ is closed under the } \omega\text{-rule.}^{33} \end{array} \right\}$.[34]

[30] I.e. $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ and $\mathcal{A}$ satisfy all the same $L_{\mathsf{P,T}}$-sentences.

[31] We still need this assumption because by assumption all $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ are $\mathbb{N}$-models, but even adding the ω-rule does not fix the standard model of arithmetic, it only fixes the *theory* of the standard model of arithmetic.

[32] I.e. there is no finite $\Delta \subseteq w$ with $\Delta \vdash^{\omega}_{\mathsf{ProbKFC}} \bot$.

[33] I.e. whenever $\{\varphi(\overline{n}) \mid n \in \mathbb{N}\} \subseteq w$ then $\forall x \varphi(x) \in w$.

[34] In fact such w are exactly the maximally $\vdash^{\omega}_{\mathsf{ProbKFC}}$-consistent set of sentences, (see Weaver, 1992, Theorem 5), but working with this characterisation is easier. That result can also be seen as a corollary of Theorems 5.12 and 5.7.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

- *For each* $w \in W_{\mathsf{c}}$, *find* $\mathbb{M}(w)$ *an* $\mathbb{N}$*-model for the language* L *such that for each sentence of* L, φ,

$$\mathbb{M}(w) \models \varphi \iff \varphi \in w.$$

- $f_{\mathsf{c}}(w) := \{n \mid \mathsf{T}\overline{n} \in w\}$.
- *For each* $w \in W_{\mathsf{c}}$, *find* $m_w : \mathcal{P}(W_{\mathsf{c}}) \to \mathbb{R}$ *a finitely additive probability measure such that for each* n,

$$m_w(\{v \in W_{\mathsf{c}} \mid \mathsf{T}\overline{n} \in v\}) = \sup\{\alpha \mid \mathsf{P}_{\geqslant}(\overline{n}, \ulcorner\alpha\urcorner) \in w\}.$$

We will show that this is well defined by showing that such $\mathbb{M}(w)$ and m_w can be found (Lemmas 5.9 and 5.11). We will also show that f_{c} is a consistent fixed point (Corollary 5.14). We finally show that the model is in fact a canonical model, i.e. $\mathsf{IndMod}_{\mathcal{M}_{\mathsf{c}}}[w, f_{\mathsf{c}}] \models \varphi \iff \varphi \in w$ (Theorem 5.12).[35]

LEMMA 5.9. *For each* $w \in W_{\mathsf{c}}$ *such an* $\mathbb{M}(w)$ *can be found.*

Proof. This could be seen as a corollary of Chang & Keisler (1990, proposition 2.2.12). It can also be proved directly as follows: Take $\mathbb{M}(w)$ to be an $\mathbb{N}$-model that interprets the contingent relation symbols by $\langle k_1, \dots k_n\rangle \in Q^{\mathbb{M}(w)}$ iff $Q(\overline{k_1}, \dots, \overline{k_n}) \in w$. Then one can prove by induction on the complexity of φ that $\mathbb{M}(w) \models \varphi \iff \varphi \in w$. □

We now present a lemma that will be useful throughout the proof.

LEMMA 5.10. *If* Δ *is* $\vdash^{\omega}_{\mathsf{ProbKFC}}$*-consistent then there is some* $w \in W_{\mathsf{c}}$ *such that* $\Delta \subseteq w$. *Therefore, if for every* $w \in W_{\mathsf{c}}$ *we have* $\varphi \in w$, *then* $\vdash^{\omega}_{\mathsf{ProbKFC}} \varphi$.

Proof. To prove this we use the Henkin method to construct some $\Delta' \supseteq \Delta$ that is finitely $\vdash^{\omega}_{\mathsf{ProbKFC}}$-consistent and "decides" each instance of the ω-rule.[36] This can then be extended to a maximally finitely $\vdash^{\omega}_{\mathsf{ProbKFC}}$-consistent set by using Lindenbaum's lemma. This will then be closed under the ω-rule because it extends Δ'. □

LEMMA 5.11. *For each* $w \in W_{\mathsf{c}}$ *such an* m_w *can be found.*

Proof. Fix $w \in W_{\mathsf{c}}$. Define $[\varphi] := \{v \in W_{\mathsf{c}} \mid \mathsf{T}\ulcorner\varphi\urcorner \in v\}$. We shall show that $\{[\varphi] \mid \varphi \in \mathsf{Sent}_{\mathsf{P},\mathsf{T}}\}$ is closed under (finite) intersection and union and contains $\varnothing$ and W_{c}, and that $\mu : \{[\varphi] \mid \varphi \in \mathsf{Sent}_{\mathsf{P},\mathsf{T}}\} \to \mathbb{R}$ by

$$\mu(\{v \mid \mathsf{T}\ulcorner\varphi\urcorner \in v\}) := \sup\{\alpha \mid \mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner\alpha\urcorner) \in w\}$$

is a monotone 2-valuation on this.[37]

The interesting case is to show monotonicity, i.e. $[\varphi] \subseteq [\psi] \implies \mu[\varphi] \leqslant \mu[\psi]$. Suppose $[\varphi] \subseteq [\psi]$. Then for each $v \in W_{\mathsf{c}}$, $\mathsf{T}\ulcorner\varphi\urcorner \to \mathsf{T}\ulcorner\psi\urcorner \in v$. Then by Lemma 5.10,

---

[35] In fact we shall show this before showing that f_{c} is a consistent fixed point because it will be used in the proof of the latter.

[36] I.e. for each $\varphi(x)$ either there is some k with $\neg\varphi(\overline{k}) \in \Delta'$ or $\forall x\varphi(x) \in \Delta'$. For details on this construction Goldblatt (2014) can be consulted.

[37] I.e. that:

- $\mu(W) = 1$
- $\mu(\varnothing) = 0$
- $A \subseteq B \implies \mu(A) \leqslant \mu(B)$
- $\mu(A) + \mu(B) = \mu(A \cap B) + \mu(A \cup B)$.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

$\vdash^{\omega}_{\mathsf{ProbKFC}} \mathsf{T}\ulcorner\varphi\urcorner \rightarrow \mathsf{T}\ulcorner\psi\urcorner$, so using rule 22 we have that $\vdash^{\omega}_{\mathsf{ProbKFC}} \forall a(\mathsf{P}_{\geqslant}(\ulcorner\varphi\urcorner, a) \rightarrow \mathsf{P}_{\geqslant}(\ulcorner\psi\urcorner, a))$. Therefore $\mu[\varphi] \leqslant \mu[\psi]$.

This μ can therefore be extended to the Boolean closure of $\{[\varphi] \mid \varphi \in \mathsf{Sent}_{\mathsf{P,T}}\}$,[38] and then extended further to m_w defined on the powerset of W_{c}.

One then needs to check that for each n this satisfies:

$$m_w(\{v \in W_{\mathsf{c}} \mid \mathsf{T}\overline{n} \in v\}) = \sup\{\alpha \mid \mathsf{P}_{\geqslant}(\overline{n}, \ulcorner\alpha\urcorner) \in w\}.$$

When $n \in \mathsf{Sent}_{\mathsf{P,T}}$ this will satisfy the equivalence by definition of μ. One needs to show that the equivalence holds for $n \notin \mathsf{Sent}_{\mathsf{P,T}}$. We can show that in that case both the left and right hand sides equal 0. For the left hand side we use the result that $\vdash_{\mathsf{KFC}} \forall x(\mathsf{T}x \rightarrow \mathsf{Sent}_{\mathsf{P,T}}(x))$[39] to show that $\{v \mid \mathsf{T}\overline{n} \in v\} = \varnothing$. For the right hand side we use axioms 3 and 20 and rule 22 to deduce that $\vdash_{\mathsf{ProbKFC}} \mathsf{P}_{\geqslant}(\overline{n}, \ulcorner 0\urcorner)$ and then use axiom 17 to get the equality. □

This $\mathcal{M}_{\mathsf{c}}$ with f_{c} is canonical, in the following sense.

THEOREM 5.12. *For every $\varphi \in L_{\mathsf{P,T}}$ and $w \in W_{\mathsf{c}}$:*

$$\mathsf{IndMod}_{\mathcal{M}_{\mathsf{c}}}[w, f_{\mathsf{c}}] \models \varphi \iff \varphi \in w$$

Proof. We work by induction on the complexity of the formula. The atomic cases mostly follow from the definition of the canonical model, although when checking the equivalence for $\mathsf{P}_{\geqslant}(t, s)$ we will also use axiom 18. For the induction step the quantifier case can be shown by the fact that $\mathsf{IndMod}_{\mathcal{M}}[w, f]$ is an $\mathbb{N}$-model and w is closed under the ω-rule. For the connective cases we use the fact that w is maximally finitely $\vdash^{\omega}_{\mathsf{ProbKFC}^-}$ consistent. □

The following lemma, along with Theorem 5.12, allows us to conclude that f_{c} is a consistent fixed point.

THEOREM 5.13. *Let $\mathcal{M}$ be a probabilistic modal structure and f an evaluation function. Then:*

$$f \textit{ is a consistent fixed point} \iff \forall w \in W(\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{KFC} \cup \mathsf{InteractionAx})$$

Proof. The direction " $\Longrightarrow$ " follows from Theorem 5.7. For the other direction we work by induction on the positive complexity of φ to show that if $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{KFC} \cup \mathsf{InteractionAx}$ then

$$\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \mathsf{T}\ulcorner\varphi\urcorner \iff (w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \varphi.$$

This and the fact that $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \forall x(\mathsf{T}x \rightarrow \mathsf{Sent}_{\mathsf{P,T}}(x))$[40] allows us to conclude that f is a fixed point. It is also consistent because of axiom 13. □

This lemma extends the useful result from Feferman that $(\mathbb{N}, S) \models \mathsf{KFC}$ iff S is a consistent fixed point. That result was generalised in Stern (2014b) where Stern shows that KFC extended by axioms for the interaction of truth with a necessity and possibility predicate, analogous to axioms 14 and 15, allows one to pick out the fixed points. Theorem 5.13 is a minor modification of Stern's result.

[38] See e.g. Zhou (2013) for a statement of this fact.

[39] This result can be found in Halbach (2014, lemma 15.16).

[40] As in Footnote 39.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

COROLLARY 5.14. *f_{c} is a consistent fixed point.*

Proof. Theorems 5.13 and 5.12 □

THEOREM 5.15. *If for every probabilistic modal structure $\mathcal{M}$ with a consistent fixed point f and $w \in W$, $\mathsf{IndMod}_{\mathcal{M}}[w, f] \models \Gamma \implies \mathsf{IndMod}_{\mathcal{M}}[w, f] \models \varphi$, then $\Gamma \vdash^{\omega}_{\mathsf{ProbKFC}} \varphi$.*

Proof. It suffices to show that every $\vdash^{\omega}_{\mathsf{ProbKFC}}$-consistent set of formulas is satisfiable. Suppose Δ is $\vdash^{\omega}_{\mathsf{ProbKFC}}$-consistent. Then by Lemma 5.10 there is some $w \in W_{\mathsf{c}}$ such that $\Delta \subseteq w$. Then by Theorem 5.12 $\mathsf{IndMod}_{\mathcal{M}_{\mathsf{c}}}[w, f_{\mathsf{c}}] \models \Delta$. Moreover we have shown in Corollary 5.14 that f_{c} is a consistent fixed point. So we have our required probabilistic modal structure, $\mathcal{M}_{\mathsf{c}}$, consistent fixed point f_{c} and $w \in W_{\mathsf{c}}$ that satisfies Δ. □

Theorem 5.5 follows from Theorems 5.15 and 5.7. The first equivalence in Corollary 5.6 is a direct corollary of Theorem 5.5, the second equivalence also uses Theorem 5.12.

§6. Conclusions. In this paper we have presented a construction of a semantics for a language that includes sentences that can talk about their own probabilities and have given a corresponding axiomatic theory. The semantics is developed by applying a familiar construction of a semantics for type-free truth, namely Kripke's construction from Kripke (1975), to possible world style structures. In this semantics some sentences are only assigned ranges of probability values instead of a single value but this will only happen for "problematic" sentences. In most cases, sentences one wants to work with will be grounded so they will then be assigned a particular probability value and one can reason in a fairly natural way. We provided an axiomatisation that allows one to reason about these semantics in a clear way. One could also use this axiomatisation to show what assumptions about probability would lead to inconsistencies.

We showed that if one expresses introspection principles by using a truth predicate to do the job of quotation and disquotation these introspection principles are consistent. Although we have only considered introspection principles here, we believe the phenomenon is quite general. For evidence of this we can see in Stern (2014a, 2014b) that the strategy worked well in the case of necessity. In future work we would like to investigate exactly how one should express principles in order to avoid the paradoxical contradictions.

One limitation of this construction is that it does not yet have the ability to account for conditional probabilities. Furthermore, it is not clear that it would be possible to add conditional probabilities and give a good definition of $(w, f) \models^{\mathsf{SKP}}_{\mathcal{M}} \mathsf{P}_{\geqslant \alpha}(\ulcorner \varphi \urcorner | \ulcorner \psi \urcorner)$ in the style of strong Kleene three valued scheme. One might overcome this limitation by instead using a supervaluational evaluation scheme. This would also result in a notion of probability that acts as an imprecise probability. Analysis of this option is left for future work.

§7. Acknowledgments. I would like to thank Hannes Leitgeb, Karl-Georg Niebergall, Stanislav Speranski, Johannes Stern and Sean Walsh and for their helpful comments on versions of this paper, as well as to two anonymous referees for this journal. I am also grateful to the audiences at the 11th Formal Epistemology Workshop at USC, Philosophy of Probability in Venice, MIRI workshop in Oxford, Philosophy of Logic Conference at UBA, Logic Tea at ILLC, Graduate Conference in Theoretical Philosophy at the University of Groningen and PhD's in Logic V at LMU. I am grateful to the Alexander von Humboldt Foundation for financially supporting my work.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

BIBLIOGRAPHY

Aumann, R. J. (1999). Interactive epistemology II: Probability. *International Journal of Game Theory*, **28**(3), 301–314.

Bacchus, F. (1990). Lp, a logic for representing and reasoning with statistical knowledge. *Computational Intelligence*, **6**(4), 209–231.

Caie, M. (2011). *Paradox and Belief*. Berkeley: University of California. Unpublished doctoral dissertation.

Caie, M. (2013). Rational probabilistic incoherence. *Philosophical Review*, **122**(4), 527–575.

Caie, M. (2014). Calibration and probabilism. *Ergo*, **1**, 13–38.

Campbell-Moore, C. (2015). Rational probabilistic incoherence? A reply to Michael Caie. *Philosophical Review*, **124**(3).

Chang, C., & Keisler, H. (1990). *Model Theory*. Amsterdam, The Netherlands: Elsevier Science. Retrieved from http://books.google.de/books?id=uiHq0EmaFp0C.

Christiano, P., Yudkowsky, E., Herresho, M., & Barasz, M. (n.d.). *Definability of Truth in Probabilistic Logic, early draft.* Retrieved from https://intelligence.org/files/DefinabilityTruthDraft.pdf (Accessed June 10, 2013).

Fagin, R., Halpern, J. Y., & Megiddo, N. (1990). A logic for reasoning about probabilities. *Information and computation*, **87**(1), 78–128.

Fischer, M., Halbach, V., Kriener, J., & Stern, J. (2015, 2). Axiomatizing semantic theories of truth? *The Review of Symbolic Logic*, *FirstView*, 1–22. Retrieved from http://journals.cambridge.org/article_S1755020314000379 doi: 10.1017/S1755020314000379

Goldblatt, R. (2014). The countable Henkin principle. In Manzano, M., Sain, I., and Alonso, E., editors. *The Life and Work of Leon Henkin*. Birkhäuser Basel: Springer, pp. 179–201.

Halbach, V. (2014). *Axiomatic Theories of Truth* (revised edition). Cambridge University Press.

Halbach, V., Leitgeb, H., & Welch, P. (2003). Possible-worlds semantics for modal notions conceived as Predicates. *Journal of Philosophical Logic*, **32**, 179–222.

Halbach, V., & Welch, P. (2009). Necessities and necessary truths: A prolegomenon to the use of modal logic in the analysis of intensional notions. *Mind*, **118**(469), 71–100.

Harsanyi, J. C. (1967). Games with incomplete information played by bayesian players, I-III part I. The basic model. *Management Science*, **14**(3), 159–182.

Heifetz, A., & Mongin, P. (2001). Probability logic for type spaces. *Games and economic behavior*, **35**(1), 31–53.

Kripke, S. (1975). Outline of a theory of truth. *The journal of philosophy*, **72**(19), 690–716.

Leitgeb, H. (2008). On the probabilistic convention T. *The Review of Symbolic Logic*, **1**(2), 218–224.

Leitgeb, H. (2012). From type-free truth to type-free probability. In Restall, G. and Russel, G., editors. *New Waves in Philosophical Logic*. New York: Palgrave Macmillan, pp. 84–94.

McGee, V. (1985). How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, **14**(4), 399–410.

Ognjanović, Z., & Rašković, M. (1996). A logic with higher order probabilities. *Publications de l'Institut Mathématique. Nouvelle Série*, **60**, 1–4.

Skyrms, B. (1980). Higher order degrees of belief. In Ramsey, F. P. and Melor, D. H., editors. *Prospects for Pragmatism.* Cambridge, UK: Cambridge University Press, pp. 109–137.

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press

Stern, J. (2014a). Modality and axiomatic theories of truth I: Friedman-Sheard. *The Review of Symbolic Logic*, **7**(2), 273–298.

Stern, J. (2014b). Modality and axiomatic theories of truth II: Kripke-Feferman. *The Review of Symbolic Logic*, **7**(2), 299–318.

Weaver, G. (1992). Unifying some modifications of the Henkin construction. *Notre Dame journal of formal logic*, **33**(3), 450–460.

Williams, J. R. G. (in press). Probability and non-classical logic. In Hitchcock, C. and Hájek, A., editors. *Oxford Handbook of Probability and Philosophy.* Oxford, UK: Oxford University Press.

Zhou, C. (2013). Belief functions on distributive lattices. *Artificial Intelligence*, **201**(0), 1–31. Retrieved from http://www.sciencedirect.com/science/article/pii/S000437021300043X, doi: http://dx.doi.org/10.1016/j.artint.2013.05.003.

MUNICH CENTER FOR MATHEMATICAL PHILOSOPHY
LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
E-mail: catrin@ccampbell-moore.com

https://doi.org/10.1017/S1755020315000118 Published online by Cambridge University Press