Available online at www.sciencedirect.com

ScienceDirect

Journal of Applied Logic 7 (2009) 177–187

JOURNAL OF APPLIED LOGIC

www.elsevier.com/locate/jal

Can logic be combined with probability? Probably

Colin Howson

Department of Philosophy, London School of Economics and Political Science, UK

Received 1 October 2007; received in revised form 11 November 2007; accepted 12 November 2007

Available online 19 February 2008

Abstract

In the nineteen sixties seminal work was done by Gaifman and then Scott and Krauss in adapting the concepts, tools and procedures of the model theory of modern logic to provide a corresponding model theory in which notions of probabilistic consistency and consequence are defined analogously to the way they are defined (semantically) in the deductive case. The present paper proposes a way of extending these ideas beyond the bounds of a formalised language, even the infinitary language of Scott and Krauss, to achieve a logic having the power and expressiveness of the modern mathematical theory of probability.

Keywords: Logic; Probability; Consistency; Formal languages

1. Introduction

Can logic be combined with probability? Trivially yes: logic can be combined with anything since it is a subtheory of any theory. Slightly less trivially, logic is necessarily involved in constraining probabilities, at least in standard formulations of probability theory. Certain deductive relationships have to be respected, for example, that if A is a logical truth then $\mathrm{P}(A) = 1$, if A and B are mutually incompatible then $\mathrm{P}(A \vee B) = \mathrm{P}(A) + \mathrm{P}(B)$,[1] and, if A and B are sentences in some language, if A is logically equivalent to B then $\mathrm{P}(A) = \mathrm{P}(B)$. This last condition is both obvious and innocent-looking, but as we shall see it is deeper than it looks.

Talk of probabilities on sentences of a language brings us naturally, and indeed inevitably, to the case of Carnap. In his monumental book [2] he inaugurated a research tradition of combining logic with probability in the following way: define a real-valued, finitely additive normalised function on the sentences of a formal language. In Carnap's case the formal language was a particularly simple one that scarcely went beyond propositional logic (monadic predicate logic without identity; like propositional logic its sentences have straightforward disjunctive normal forms). Carnap called the result 'logical probability', without much in the way of argument to support this terminology, and none of it very convincing. If one wanted to be funny one might describe Carnap's model as affirming something like

$$\text{logic} + \text{probability} = \text{logical probability} \sim \text{cheese} + \text{soufflé} = \text{cheese soufflé}$$

E-mail address: c.howson@lse.ac.uk.

[1] As usual '$\vee$' is the symbol for inclusive 'or'.

1570-8683/$ – see front matter
doi:10.1016/j.jal.2007.11.003

A quite different model is that in which probability theory and logic are not seen as conceptually disparate disciplines whose sole point of contact is a formal language, but as complementary logics in their own right, with a common set of concepts and methods. In what follows I will describe some attempts to develop a theory along these lines, and suggest a way to develop its full potential.

2. Probabilistic model theory

It is difficult not to be struck by the fact that the formalisms of logic and probability are remarkably similar, at any rate in their semantic aspects. Both are domain-general theories of validity, consisting of a type of propositional algebra A and a real-valued non-negative additive valuation function V on A. In deductive logic V is of course a truth-valuation, taking values in the set $\{0, 1\}$ (as usual 1 will represent 'true' and 0 'false'). The question is whether this formal similarity is simply that, a similarity, or whether it should be interpreted as suggesting a close degree of conceptual kinship.

De Finetti, at least in his earlier writings, seems to have regarded the rules of epistemic probability as an authentic species of logic. Thus not only is the title of his famous 1937 paper 'La prévision: *ses lois logiques*, ses sources subjectives' [6] (my emphasis) but also, in a paper written shortly before, he wrote that

> It is beyond doubt that probability theory can be considered as a multi-valued logic (precisely: with a continuous range of values), and that this point of view is the most suitable to clarify the foundational aspects of the notion and the logic of probability. (De Finetti [5], parenthesis in the original; quoted in [3, p. 61]. These values are not additional truth-values, but probability-values 'superimposed' [sic] on the logic of truth-values.)

He went on to claim that the rules of subjective probability are *objective consistency constraints* on probability-evaluations,[2] just as the truth-valuation rules are consistency-constraints for truth-value assignments (Smullyan [26] gives a proof-theory for the latter in terms of 'signed' semantic tableaux; I shall mention this interesting parallel again later).

These observations raise an obvious question. The semantic characteristic of (deductive) consistency is possession of a model, and the question is this: Do 'salient' features of deductive model theory extend naturally to probability? In a seminal paper published in 1964 a young co-worker of Carnap's, Haim Gaifman [9], went a long way to answering this question positively. Noting that a relational structure for a first order language L can be represented as a pair (D, m), where m is an additive function assigning values in $\{0, 1\}$ to atomic sentences of $L(D)$ ($L(D)$ is L plus enough constants to name all members of D), extended to a valuation of all sentences in $L(D)$ via a standard recursive definition, Gaifman generalised this to a *probability model*, of the form (D, m) where m is a finitely additive function taking values in $[0, 1]$. A technical problem is that a probability on the atomic sentences does not determine a probability on all the sentences of $L(D)$. However, Gaifman proved that a probability on the *quantifier-free sentences* of $L(D)$ has a unique extension to all the sentences of $L(D)$ satisfying what has since come to be called the *Gaifman Condition*, namely:

$$m\big(\exists x A(x)\big) = \sup m\big(A(a_{i1}) \vee \cdots \vee A(a_{ik})\big),$$

where the sup is over all finite sets of constants of $L(D)$. This result is clearly a 'logical' analogue of the extension theorem for countably additive measures on Boolean σ-algebras. Where $\mathrm{Mod}(A)$ is the set of all models of a sentence A of L, the set S of all such sets is a field of subsets of $\mathrm{Mod}(T)$ where T is a logical truth of L. It is a well-known consequence of first-order compactness that the induced probability m^* is continuous on S and so has a unique countably additive extension to $B(S)$, the Borel field generated from S. It is not difficult to show that, if $\mathrm{Mod}(T)$ is restricted to countable structures, then $m^*(\mathrm{Mod}(\exists x A(x))) = \sup m^*(\mathrm{Mod}(A(a_{i1}) \vee \cdots \vee A(a_{ik})))$, so Gaifman's condition transferred to S is guaranteed by the Carathéodory extension theorem for countably additive measures.

Gaifman's paper inaugurated a new branch of logic 'continuously' extending the model theory of first order logic (it becomes a limiting case). Shortly afterwards Dana Scott and Peter Krauss [22] extended Gaifman's work in two

[2] It would take us too far afield to go into the reasons why 'coerenza', which would be naturally translated as 'consistency' and in some of de Finetti's published work actually is, eventually became 'coherence'.

important ways: they (i) considered language systems closer in expressive power to the σ-algebras of mathematical probability, and (ii) developed a corresponding model theory and proof theory. Just as with 'ordinary' logic, however, we shall see (i) and (ii) pull in opposite directions.

2.1. Expressive power

The typical structure investigated in mathematical probability is a σ-algebra with a countably additive probability function defined on it. A first order language L determines only an ordinary Boolean algebra (via the map from sentences A to either $\mathrm{Mod}(A)$, the class of models of A, or to the corresponding member $|A|$ of the Lindenbaum sentence algebra of L), not closed under countable sup and inf. However, there is a well-known extension of first order logic, extending an ordinary first order language L, obtained by closing off its class of formulas under the formation of countable conjunctions and disjunctions but retaining finite strings of quantifiers. This is the *infinitary language* of type $L_{\omega 1,\omega}$ in the $L_{\kappa,\lambda}$ family, where κ and λ are infinite cardinal numbers and $L_{\kappa,\lambda}$ is a first order language closed under the formation of fewer than κ conjunctions and disjunctions, and strings of quantifiers of length less than λ. What singles it out from all the other languages in this family for $\kappa \geqslant \omega_1$ is that, despite its large increase in expressive power over first order languages, i.e. over $L_{\omega,\omega}$,[3] and the fact that because of introduction and elimination rules for the countable $\wedge$ and $\vee$ operations proofs may be of countably infinite length, there is an effectively-specifiable proof system with the property that every semantic consequence of a countable set of sentences is a provable consequence (though this does not extend to consequences of uncountable sets).

Since the Lindenbaum algebra of $L_{\omega 1,\omega}$ is isomorphic to a σ-field of sets, of the form $\mathrm{Mod}(\varphi)$ for sentences φ in $L_{\omega 1,\omega}$, many of the characteristic structures of mathematical probability theory are definable in $L_{\omega 1,\omega}$. Suppose, for example, that the sample space is $C = 2^{\omega}$, a set that in descriptive set theory is often identified with the real numbers. If the base first order language for $L_{\omega 1,\omega}$ contain a denumerably infinite set of constants a_i and a one-place predicate symbol $B(x)$, the members of C can be identified with the structures (up to isomorphism) of the appropriate type in $\mathrm{Mod}(\sigma)$ where $\sigma = \bigwedge_{j<\omega} \bigwedge_{i<j} \neg(a_i = a_j) \wedge \forall x[\bigvee_{i<\omega}(x = a_i)]$. The basic neighbourhoods are definable by ordinary first order sentences conjoined with σ, and hence the Borel sets in C are definable in $L_{\omega 1,\omega}$:

> ... $L_{\omega 1,\omega 0}$ is the proper generalisation of $L_{\omega 0,\omega 0}$ to denumerable formulas. In fact we have seen several reasons for claiming that $L_{\omega 1,\omega 0}$ plays the same role for $L_{\omega 0,\omega 0}$ that the theory of Borel sets and σ-fields plays for the ordinary field of sets' [21, p. 341].

2.2. Model theory and proof theory

Scott and Krauss extended Gaifman's notion of a probability model to $L_{\omega 1,\omega}$, as an ordered set whose members are a complete Boolean algebra B, a set D with a relation Id mapping $D \times D$ into B (with $\mathrm{Id}(a,a) = 1$ for all a in D), and a strictly positive countably additive probability function m on B (B is complete because it interprets quantifiers as sups and infs over all the instances of the quantified formulas specifiable within the domain). If m is two-valued B is the Boolean algebra 2 and the model reduces to a classical relational structure.

It remains to specify what these structures are going to be models *of*. Because in classical logic there are only two truth-values they do not have to be, and usually are not, explicitly represented in the primitive vocabulary (negation suffices). But it can be and has been done: Smullyan [26] developed a proof theory for first order languages with truth-valued, or in his terminology 'signed', sentences, having the form AT and AF, which we can suggestively represent $v(A) = 1$, $v(A) = 0$.[4] Things are more complicated in the probabilistic case because real-valued probability assignments should be able to specify a variety of relationships among real numbers; e.g., as taking values within certain intervals, and more generally having properties and relations specifiable in some suitable mathematical language/theory T. Let us leave T as a free parameter for the moment. These probability assignments, called 'probability assertions' by Scott and Krauss, are defined formally as sequences $(\Phi, \mathbf{A})$ where Φ is a for-

[3] For example, a single $L_{\omega 1,\omega}$ sentence defines the standard model of arithmetic, and another the class of finite sets.

[4] For fairly obvious reasons no embedding is permitted: despite what looks like a blurring of the object-language/metalanguage distinction, this is a calculus of truth-value assignments to object-language sentences in a semantically open language.

mula of $L(T)$ with n free variables and $\mathbf{A} = (A_1, \ldots, A_n)$ is a finite sequence of sentences of $L_{\omega 1,\omega}$. Intuitively $(\Phi, \mathbf{A})$ 'says' that the probabilities of $A_1, \ldots, A_n$ satisfy Φ. So (Tarski's T-schema!) $(\Phi, \mathbf{A})$ is true in a probability model with probability function m just in case $\Phi[m(A_1), \ldots, m(A_n)]$ holds in the intended model of T. 'Valid probability assertion' and 'valid consequence' are defined analogously to the deductive case: where α is a probability assertion and Σ a set of assertions 'α is valid' = 'α is true in all probability models' and 'α is a consequence of Σ' = 'every probability model of Σ is a model of α'.

For the choice of T Scott and Krauss, like several people since, selected the first order theory M of real closed fields. This is algebraic real number theory, in which among other things inequalities and equalities between polynomials with real coefficients can be expressed and evaluated. A famous result of Tarski is that M admits quantifier-elimination and so is decidable (and complete). The existence of a proof procedure for $L_{\omega 1,\omega}$-logic and the decidability of M were exploited by Scott and Krauss in demonstrating that the set of valid probability assertions is effectively enumerable ([22, Theorem 6.7]: this is not proved in the usual way via some axiomatisation; instead Scott and Krauss use the familiar fact that any probability statement can be put in an equivalent normal form involving only propositional atoms, and then show that there is an algorithm for enumerating the valid ones).

Despite this being a remarkable result, I think that the qualifications that need to be made remove a good deal of its attractiveness. In particular, the natural coding structure for the syntax and proof theory of $L_{\omega 1,\omega}$ is the set HC of hereditarily countable sets (sets of countable sets whose members are countable, etc.). But HC is a very much larger set than the coding structure for countable first order languages, which is N if you use ordinary Gödel numbering or the class HF of hereditarily finite sets (i.e., the set $R(\omega)$ of sets of finite rank in the cumulative hierarchy): HC also has a number-coding, but only in the *real numbers*. Hence 'effectively enumerable' in the context of $L_{\omega 1,\omega}$-validity certainly does not mean recursively enumerable.[5] So what does it mean? It means this. A formula in the language of first order set theory is said to be Σ_1 (Π_1) if it is built up from atomic formulas $x \in y$ using only the connectives, bounded quantifiers $\exists x \in w$ ($\forall x \in w$), and unbounded $\exists x$ ($\forall x$). 'Effectively enumerable' in the discussion above just means 'definable in HC by a Σ_1 formula', by analogy with 'recursively enumerable' in the sense of ordinary recursion theory, viz. 'definable in HF by a Σ_1 formula' (equivalently, 'definable in N by a relation').

But being a Σ_1-definable subset of HC() will not strike many as providing an intuitively acceptable surrogate of 'effectively enumerable', for it translates into being a Σ^1_2 subset of the *reals*. This is certainly not the meaning of 'effectively enumerable' in the sense of being in the output of a Turing machine (i.e., a Σ^0_1 subset of N). Nor does one achieve that 'gold standard' of effectivity by intersecting the set of HC-codes of $L_{\omega 1,\omega}$ formulas with a certain type of *countable* set A (technically a *countable admissible set*) which supports a more authentic generalisation of recursion theory, even though the result enables more of the attractive features of first order logic to be reproduced, *mutatis mutandis*; notably, as Barwise proved in a celebrated result, weak compactness. The notions of A-recursiveness (Δ_0-definable in A) and A-recursive enumerability (Σ_1-definable in A) remain still far stronger than ordinary recursiveness and recursive enumerability for any admissible set: even with respect to the smallest countable admissible set, $L(\omega_1^{\mathrm{CK}})$,[6] which does not collapse $L_{\omega 1,\omega}$ into $L_{\omega,\omega}$ the corresponding sets still lie in the analytic hierarchy.[7]

So for most people accustomed to a notion of logical reasoning as one which when formalised issues in a recursively enumerable set of theorems, the proof theory for probability assertions based on $L_{\omega 1,\omega}$ goes a long way beyond their notion of what a logic is or should be. On the other hand, as far as capturing ordinary probability theory is concerned, it is still relatively weak. For example, because their probability assertions have only finitely many members it is not a valid assertion that the probabilities of a countable set of mutually inconsistent sentences whose disjunction is a logical truth sum to 1, which is ironical since the point of having countable operations available in the language was precisely to be able to say and prove such things. A more adequate proof theory for probability assertions would however be even more distant from what is ordinarily regarded as logic. Gaifman and Snir [10] show that a surprising amount of the mathematics of probability can be done using a base language consisting of the first order language $L(N)$ of arithmetic augmented by a supply of additional function and predicate symbols. They also obtain some novel results: for example, by Gödel numbering the formulas and using well-known techniques for obtaining rational

[5] If $L_{\omega 1,\omega}$ is replaced by an ordinary first order language then the set of valid probability assertions is recursively enumerable [22, Theorem 7.6].

[6] $L(\omega_1^{\mathrm{CK}})$ is the set of all sets constructible, in Gödel's sense, before the first nonrecursive ordinal (CK stands for 'Church–Kleene').

[7] The $L(\omega_1^{\mathrm{CK}})$-recursively enumerable sets are Π^1_1 subsets of N, and the $L(\omega_1^{\mathrm{CK}})$-recursive sets are Δ^1_1 (hyperarithmetical); among the latter is the set of Gödel numbers of sentences true in the standard model of arithmetic [16, pp. 15–17].

approximations to real-valued functions, they are able to show where various classes of probability functions definable in $L(N)$ appear in the arithmetical hierarchy.

But first order languages are nevertheless a very Procrustean bed into which to fit the sort of reasoning in informal probability theory which may and does involve consideration not just of real numbers but sets of real numbers, sets of infinite sequences of reals, and so forth. The classical logic in which reals and sets of reals are simultaneously discussible is of course second order (second order arithmetic). Despite some recent strong advocacy however [25], second order logic is regarded by many as too distant from being an acceptable logic for two reasons. One is that, with the standard semantics, it is indistinguishable from set theory, and controversial set theory at that: there are sentences of pure second order logic which are valid (true in all interpretations) just in case, respectively, the axiom of choice is true, the generalised continuum hypothesis is true, there is a strongly accessible cardinal, etc. The other reason is that because of its non-axiomatisability with respect to standard semantics second order logic is very far from the effectiveness criterion that many people regard as essential to an organon of reasoning.[8] There is of course a formulation of second-order logic that is axiomatisable, for a semantics where the domain of the second-order predicate variables need not be the full power set of that of the first. But (a) this move deprives second-order formulations of its characteristic ability to characterise structures up to isomorphism, or 'near-isomorphism' in the case of second-order set theory, and more importantly (b) just as relativising $L_{\omega1,\omega}$ to admissible sets deprives it of a uniform interpretation, so Henkin semantics for second-order logic deprives the second-order quantifiers of a uniform interpretation.

3. A modest proposal [9]

We seem to have reached an impasse. The desideratum is an interpretation of the formalism of Bayesian probability as a logic in which all, or at any rate most, of the reasoning carried out in 'informal' mathematical probability can be reproduced. In the light of the foregoing remarks about the contrary pulls of expressiveness and effectiveness, this would seem to be impossible. Or is it? I shall now propose what I think is a reasonable resolution of the problem, but first let us review some familiar facts. (i) The mathematical theory within which probabilistic reasoning is standardly carried out is set theory, usually informal set theory, augmented by Kolmogorov's axioms. Set theory is a universal language and theory for mathematics. (ii) Set theory is axiomatisable in a countable, in fact finite, first order language, and its theorems, possibly augmented by additional assumptions (e.g., those describing a particular probability space) are therefore a recursively enumerable set. (iii) The possibility space S in the usual Kolmogorov formalism can be interpreted as either a space of generic outcomes of a stochastic experiment, or as the set of logical possibilities within which a σ-field of *propositions*, determined by the problem at hand, is defined (it is almost commonplace in discussions of epistemic probability for the subsets of S to be referred to as propositions; recall the earlier observation about the similarity of the formalisms of orthodox mathematical probability and logic).

I would now like to make a modest proposal, namely to take as our object-languages *not formal languages at all* but set-theoretic algebras, including σ-algebras, as algebras of *propositions* (there is no reason in principle not to disregard the set-theoretic basis at all and simply take the propositions as elements of some Boolean σ-algebra, but by Stone's theorem the set-theoretical formulation is an equivalent formulation, and is also the usual one for informal probability).[10] As I noted a little earlier, characterising the class of probability-bearing propositions as subsets of a logical possibility space is already a feature of many if not most discussions of epistemic probability (so many, in fact, that it would be tedious and space-wasting to provide references). Moreover, given that if $A \Leftrightarrow B$ then $\mathrm{P}(A) = \mathrm{P}(B)$, the propositional as opposed to sentential formulation is virtually forced on us as being more fundamental. Gaifman and Snir's observation that it can be highly informative to use formal languages, e.g., in the classification of degrees of dogmatism of probability functions in terms of the arithmetical hierarchy, is not overlooked, since it is effectively subsumed via the map from sentences A to the subsets $\mathrm{Mod}(A)$ of the space represented by $\mathrm{Mod}(T)$ for T a logical truth of the language. So nothing is lost from the formal language approach.

[8] The set of Gödel numbers of valid second-order formulas is not even in the analytical hierarchy (i.e., is not second-order definable).

[9] Cf. [28].

[10] Characterising statements simply as elements of an appropriate *algebraic structure* rather than as explicitly linguistic entities is not at all unusual these days (we can ignore the fact that formulas of a formalised language are themselves members of a type of freely-generated algebra). For example, most discussions of quantum logic identify it as the projection lattice of a Hilbert space.

The basic semantics for characterising the consistency etc of probability assignments to elements of such algebras are very similar to those of Scott and Krauss for their probability assertions. To keep things simple let us just consider assignments of real-valued probabilities to elements of a field or σ-field F of sets. Let us call a real-valued assignment $V[E]$ to a subset E of F (E is a set of *propositions*) *consistent* on F if it can be extended to F in a way that satisfies the constraints imposed by the finitely additive probability axioms (cf. the corresponding definition for a propositional language [26, p. 10]). Equivalently, $V[E]$ is consistent if it is the restriction of a probability function on F to E. The following theorems are clearly analogues of corresponding (meta)results for first order logic:

Theorem 1 *(Absoluteness of consistency). If E is a subset of a field F and $V[E]$ is consistent on F then it is consistent on any field F' including F (and so we can just talk of consistency* tout court; *note that this would not be true if countable additivity had been assumed).*

Proof. The result, which is actually a corollary of the Hahn–Banach theorem, follows immediately from two well-known results of de Finetti. Say that a real-valued assignment $V[E]$ to the members of E has the property $\Pi(V)$ just in case there is no finite subset $E = \{E_1, \dots, E_k\}$ of F and no numbers $q_1, \dots, q_k$ such that every sum X of bets on members of E at betting quotients $V(E_1), \dots, V(E_k)$ and stakes $q_1, \dots, q_k$ is not positive (or negative: reverse the signs of the q_i) whatever the truth-values of the E_i. The first of these results, often called the *Dutch Book theorem*, states that if V is a real valued function on a field F of subsets of a set S such that $V(S) = 1$, then V is finitely additive on F iff it has $\Pi(V)$. De Finetti's second result [7, p. 78] states that $V[E]$ can be extended to any field which includes E iff V has the property $\Pi(V)$. □

Define a *model* of $V[E]$ to be an extension of $V[E]$ to F satisfying the probability constraints (cf. the definition of a model of a truth-value assignment to a set of sentences in a propositional language [26, p. 9]). Consequence is defined analogously to semantic consequence in classical logic: where A is a set of propositions then $V(A)$ is a *consequence* of $V[E]$ if every model of $V[E]$ is a model of $V(A)$. By analogy again with deductive logic we can define the *content* of an assignment $V[E]$ to be the set of finitely additive probability functions not satisfying $V[E]$, and we thereby obtain:

Theorem 2 *(Non-ampliativity of consequence). If $V(A)$ is a consequence of $V[E]$ then the content of $V(A)$ does not exceed that of $V[E]$.*

Proof. Immediate. This is of course the analogue of the fundamental 'nothing in, nothing out' result supposedly characterising deductive consequence and widely thought, as we now see erroneously, to be violated by probabilistic inference. □

Theorem 3 *(Compactness). $V[E]$ is consistent iff $V[E_0]$ is consistent for every finite subset E_0 of E.*

Proof. Suppose $V[E]$ is inconsistent. It follows from the second of the de Finetti theorems mentioned in Theorem 1 that $V[E]$ does not have the property $\Pi(V)$. Hence neither does $V[E_0]$, where E_0 is a finite subset of E. Hence, by the same theorem, for some finite subset E_0 of E, $V[E_0]$ is inconsistent. □

This is all very natural, and by itself I believe fully justifies the title of 'logic'. But what about a proof-theory? Since we are no longer in the realm of formal languages it might seem that we are in no position to expect one, at any rate a formal proof theory of the sort we associate with first order logic and various other effective logics. What proof theory we do have is supplied by Kolmogorov's axioms or (if you prefer to take conditional probability as primitive while remaining in a set-theoretical framework, Renyí's) together with the general set theory axioms. It is within this familiar framework that we deduce that certain assignments are inconsistent, that this one is a consequence of those, and so forth. Moreover one can formalise any piece of set-theoretical reasoning within first order set theory together with additional assumptions characterising the features of the problem at hand; and the theorems are, of course, a recursively enumerable set.

It might be objected that this rather quick way of uniting the almost unlimited expressibility of set theory with the gold-standard of effectiveness, recursive axiomatisation, is just a trick, and one in particular which ignores the

standard objection that first order set theory, like any first order theory with an infinite model (assuming it has one at all) is not categorical: there will always be non-standard models.[11] While that is of course true, it is not clear why non-categoricity should be thought to pose a problem. We *know* what the standard model of set theory is: we can even refer *univocally* to it, or at any rate to a big enough initial segment of it, in a second order formalisation. You simply cannot get a *complete* theory of that model, which is why second order *logic* itself must be incomplete. But the incompleteness is endemic, whether in first or second order logic, with the incompleteness of second order logic having its counterpart in the incompleteness of first order *theories*, including first order set theory. Indeed, in terms of what each can do there is virtually no difference between first order set theory and second order logic (cf. [29, p. 19]).

4. The countable additivity question

Theorems 1 and 3 above depend on not admitting the rule of countable additivity among the probability axioms. With probabilities defined on finitary languages the issue of course does not arise. But on σ-algebras it definitely does, and must be answered.[12] The answer becomes particularly pressing in the context of the present discussion, for the results obtained above hold only under finite additivity. The question preoccupied de Finetti, and he famously rejected countable additivity as a universal axiom. His reasons are in some tension with his celebrated operationalism because, as we shall see, it is easy to construct examples where non-countably additive distributions can be Dutch-Booked, i.e. where, if those probabilities are regarded as fair betting quotients, stakes can be set such that the bettor at those odds will suffer a certain loss. So let us briefly review his case against countable additivity.

The common argument, used by among other people Kolmogorov himself, that countable additivity is justified by the technical convenience it undoubtedly procures de Finetti dismissed with the plausible observation that mathematics should be servant rather than master, and in particular the servant of reflective intuition about what should and what should not be regarded as determining mere consistency. He illustrated the workings of this intuition as it bears on the additivity problem with some informal examples. One of these concerns the possibility of a countably infinite fair lottery. Under countable additivity is of course not a possibility. But it seems odd that a criterion of *consistency* should forbid a uniform distribution over a countable disjoint set, as is certainly possible with either a finite one or a bounded interval of real numbers, and instead demand, as it does, a heavily skewed one. An even more compelling example (I think) is this: if you first assign a uniform distribution to a variate ranging over the unit interval, and then receive information simply that its value is a rational number, it seems highly implausible to regard this information *by itself* as implying that the corrected distribution over the countable set of rationals must be heavily skewed. Here countable additivity seems to be *adding new information* not contained in the conditioning information, namely that some of the probabilities are now much more likely than others. De Finetti as usual puts the point succinctly: 'Here the *content* of my judgment enters into the picture' ([8, p. 123], emphasis in the original).

He might have made the same general comment about the well-known Bayesian convergence-of-opinion theorems, which in their strong 'with probability 1' formulation require countable additivity. Since for a countably additive distribution over a countable partition a finite subset will carry a probability $1 - \varepsilon$ where ε tends to zero, this implies that if a hypothesis H about a data source generating countably infinite data sequences is false the probability that it will be falsified after any given finite number of observations must tend to 0. It follows that sufficient positive evidence will push the probability of H arbitrarily close to 1 and we seem to have a solution to the problem of induction out of pure mathematics. Kelly makes this point explicitly [17, pp. 321–330].

Another objection to countable additivity, which de Finetti mentions but does not stress because he famously did not believe in objective probability, is that long-run relative frequencies are not generally countably additive. But for those who are prepared to admit long-run frequencies as measures of objective probabilities and wish to base subjective probabilities on them where they are known,[13] this is a powerful argument. Though some think there are

[11] Skolem's Paradox is probably the most famous example: by the Löwenheim–Skolem theorem if ZFC has a model it has a countable sub-model M. But by Cantor's theorem, which is of course true in M, the power set of ω is uncountable.

[12] Many people simply assume that probabilities *must* be countably additive. Cf. Halpern: 'For one thing there are no uniform probability functions in countable domains. (Such a probability function would have to assign probability 0 to each individual element in the domain, which means by countable additivity it would have to assign probability 0 to the whole domain.)' [12, p. 315]

[13] The principle that subjective probabilities P should be set equal to objective chances P* where the latter are known is usually expressed in the following conditional form $\mathrm{P}(A|\mathrm{P}^*(A) = p) = p$. Assuming that the possibility space S contains the possible values of an assumed objective

methodological problems with characterising probabilities as limiting relative frequencies, there is no doubt that there exists a consistent mathematical theory of them, namely von Mises's theory of Collectives. It is also easy to find models in that theory of the uniform distribution over a countable partition in that theory: e.g., let the Collective be any permutation of the set N of natural numbers and the attributes the singletons $\{n\}$, $n = 0, 1, 2, \ldots$. The limiting relative frequency of each exists and is equal to 0, though the countable union (disjunction) has of course limiting relative frequency 1 (von Mises's axiom of randomness, in Church's recursion-theoretic form, is also satisfied, if rather trivially). It is well known that events with well-defined limiting relative frequencies do not always form a field, but Kadane and O'Hagan [14] show that the uniform distribution over the singletons of N is extendable to all subsets of N.

This is not to say that there are not also arguments in favour of countable additivity. Probably the most important, as far as the literature is concerned at any rate, is the fact that non-countably additive probability functions whose domains include countably infinite partitions are *nonconglomerable* with respect to at least one such partition, and conversely. A probability P is conglomerable with respect to a countable partition $B = \{B_i\colon\ i = 1, 2, \ldots\}$, if for every proposition A in the domain of P and numbers x, y such that $x \leqslant \mathrm{P}(A|B_i) \leqslant y$ for all B_i in B, $\mathrm{P}(A)$ lies within the same bounds; nonconglomerability is just the negation of conglomerability. I shall follow the usual custom and truncate 'conglomerable/nonconglomerable with respect to countable partitions' to just 'conglomerable/nonconglomerable'. Nonconglomerability sounds rather like a failure of the infinitary logical rule of 'or'-introduction. The appearance is illusory, however, for that would require parsing a conditional probability $\mathrm{P}(A|B)$ as 'the probability of A if B is true', a parsing which as de Finetti shows is untenable [7, p. 104].

There is no denying though that there are some odd features of nonconglomerability. Consider again the fair countable lottery, and the following scenario described, de Finetti tells us [7, p. 205], by Lester Dubins in a letter to Savage. Two different mechanisms, A and B, for randomly generating a positive integer N are each selected with probability 1/2, and we are given that $\mathrm{P}(N = n|A) = 2^{-n}$, while $\mathrm{P}(N = n|B) = 0$ for all n (thus violating countable additivity). It follows that $\mathrm{P}(B|N = n) = 0$, for all n, though $\mathrm{P}(B) = 1/2$: P is nonconglomerable in the partition $\{\{n\}, n = 1, 2, 3, \ldots\}$. Note that in this example all the conditional probabilities are fully determinate, and that we also have, in the limiting relative frequency model, a concrete model of a random distribution over N. But now we seem to have a paradox, because the identity $\mathrm{P}(N = n|B) = 0$ for all n tells us that no individual value of N conveys any discriminatory information, yet one should nevertheless bet at infinite odds on A and against B after any given observation even though their unconditional probabilities are 1/2! In other words, you know in advance of making any observation that *whatever its outcome* it will decide you with certainty in favour of A, yet even armed with that foreknowledge you still only assign A a priori probability of 1/2. As Kadane et al. [15] neatly put it, with any nonconglomerable probability function you are always liable to 'reason to a foregone conclusion' (whereas such reasoning is always precluded by countable additivity). Worse, you seem to be reasoning inconsistently: since you know that '$N = n$' must be true for some n, it looks as if you are in effect assigning different probabilities to the same proposition (both to A and to B).

While that claim is of course not strictly speaking true, there is another, and some say related (I shall return to this claim shortly), pathology in the probability function above: if you were to bet at the odds determined by it you would be vulnerable to a *Dutch Book*, i.e. to the existence of a set of stakes which would make you certain to lose come what may. In fact, *any* nonconglomerable probability—and hence any non-countably additive function[14]—is vulnerable in the same way. Suppose that $\mathrm{P}(C|D_i) \leqslant k < 1$ and $\mathrm{P}(C) > k$, where D_i, $i = 1, 2, \ldots$, is a countable partition. Let W be a bet paying \$1 if C is true and $W|D_i$ be the same bet conditional on D_i. Suppose that, as usual, conditional preferences are represented by conditional expectations and unconditional preferences by unconditional expectations. Then a payment of \$$k$ is weakly preferred to W conditional on B_i, for each i, whereas W is preferred to a payment of \$$k + d$ for some $d > 0$. This violation of countable dominance is easily seen to be a Dutch Book. For if you own P you will regard selling a conditional bet on C given D_i with stake \$1 and betting quotient b as at worst fair, for each i, and buying a bet on C with betting quotient $k + d$, for some $d > 0$, and stake \$1 as more than fair. Since one of the D_i must be true, the net gain from all these bets is \$–d.

probability, the conditioning statement is not a 'first order probability', as some claim, but merely a statement of the form $X = p$ where X is an appropriate random variable on S.

[14] See [31].

The question is, however, of what we should see vulnerability to a Dutch Book as revealing about the underlying evaluations. A thorough consideration of this question, a controversial one in the Bayesian literature, is far too big an issue to go into here, and I shall content myself here merely with a few remarks. As de Finetti himself pointed out, Dutch Book arguments generally depend on the assumption, call it (*), that the sum of any n bets is fair just in case each is fair [7, p. 77], and he famously went on to deny that (*) can be extended to infinite sums of bets, even where these are defined, precisely because he believed that (*) fails to hold in the infinite case. Indeed, he explicitly regarded the uniform distribution, uniformly 0, in a countably infinite lottery as perfectly consistent even given its obvious vulnerability to a Dutch Book [7, p. 91]. But as Schick was probably the first to point out, (*) is just as questionable, *and may well be false*, for finite as well as infinite sums of bets [24].

Despite this, many people still believe that Dutch Bookability is a symptom of genuine inconsistency, that of evaluating uncertainties (propositions, bets, uncertain options generally) differently depending on how they are expressed or described, and they use this claim to argue that the familiar Dutch Book arguments for the probability axioms do no more than exhibit such an inconsistency in violations of those axioms. Ramsey famously expressed this view in a much quoted passage (1926, p. 80) and Skyrms, a modern follower (quoting that passage), illustrates it with the example of the law of binary additivity. If what I said in the preceding paragraph is correct then Skyrms's argument must be erroneous. Let us see. Suppose then that $W(A, p, r)$, $W(B, q, r)$ are bets on A and B, respectively, with the same stake r and probabilities, considered as fair betting quotients, p and q, where A and B are mutually inconsistent. The agent can certainly be Dutch Booked in simultaneous bets by giving odds determined by p and q on A and B and odds not equal to $p + q/(1 - (p + q))$ on $A \vee B$. Hence, according to Skyrms,

> if you are to be *consistent*, your personal probability for $[A \vee B]$ had better be ... probability$[A]$ + probability$[B]$. ([27, p. 21]; Skyrms uses p and q for my A and B.)

But that does not follow. To be consistent in the Skyrms–Ramsey sense merely means acknowledging that $W(A, p, r) + W(B, q, r) = W(A \vee B, p + q, r)$, which is perfectly consistent with *denying* that the sum of two fair bets on A and B is the fair bet on $A \vee B$, i.e. denying (*). What is being confused here is *being a functional*, i.e. an evaluation independent of the precise form in which its subject is described, and *being a linear functional*, i.e. one whose value on a sum is the sum of its values on the summands—in effect, a functional satisfying (*). To be fair to Skyrms, he implicitly concedes this in a footnote by pointing out that the argument in his text does depend on assuming that the expected value of a sum of two bets is the sum of their expected values; but this is of course (*) [27, p. 123, n. 4].

5. The axioms of probability

We have seen earlier de Finetti argue that a purely 'formal' principle should not forbid *in principle* a uniform distribution over the elementary events (atoms) in a power set algebra. It seems to me that this is so fundamental a principle that it should count as a further probability axiom (note that it is *not* an endorsement of the Principle of Indifference, which is prescriptive rather than permissive). In addition there is the fact that we get a nicely analogical development of a logical probability without countable additivity: neither compactness nor the extendability of a consistent assignment of probability to an arbitrary including algebra hold under countable additivity. It would be a very nice bonus to have a type of completeness result that told us that the laws of probability should include finite but not countable additivity. Were that so then we would have a powerful endorsement of de Finetti's claim that countable additivity should count as no more than a special assumption to be used as the problem at hand is thought to demand.

Dutch Book arguments certainly do not have this character because, as we have observed, they would include rather than exclude countable additivity, and neither does the usual utility-based one, whose classic exposition is Savage [20], since a continuity condition can be straightforwardly added to give countable additivity [30].[15] Neither do the stand-alone sets of axioms for a qualitative probability ordering that have been investigated: for a probability

[15] There are of course also difficulties with a Savage-type account: not only the familiar utility paradoxes but also the deeper problem of unequivocally separating out utilities from probabilities described in [23].

function to agree, a qualitative additivity principle has to be appended, which begs the question, and in addition continuity assumptions can be added to give countable additivity. An alternative approach is to start immediately with a quantitative notion and think of general principles that *any* acceptable numerical measure of uncertainty should obey. R.T. Cox [4] and I.J. Good [11], working independently in the mid nineteen-forties, showed how strikingly little in the way of constraints on a numerical measure yield the finitely additive probability functions as canonical representations. It is not just the generality of the assumptions that makes the Cox–Good result so significant: unlike some of those which have to be imposed on a qualitative probability ordering, the assumptions used by Cox and to a somewhat lesser extent Good seem to have the property of being *uniformly* self-evidently analytic principles of numerical epistemic probability *whatever particular scale it might be measured in*.[16] Moreover, these principles do not extend to endorsing countable additivity.

Cox, whose treatment rather than Good's will be the one I shall refer to in what follows, identified three such invariant principles. Let M be an admissible real-valued *conditional* measure (since one of the objectives is to generate the multiplication principle), taking values in an interval of the real line. Then for any jointly consistent B, C, D

(i) $M(\neg A|C) = f(M(A|C))$, for some real-valued, twice-differentiable function $f(x)$ decreasing in x.
(ii) $M(A \,\&\, B|C) = g(M(A|B \,\&\, C), M(B|C))$, for some real-valued $g(x, y)$ with continuous partial derivatives, increasing in x and y.[17]
(iii) If $A \Leftrightarrow B$ and $C \Leftrightarrow D$ then $M(A|C) = M(B|D)$.

Cox sets out at some length the reasons why these three rules deserve to be regarded fundamental, and I shall simply refer the reader to his exposition [4]. His next, and major, step was to show that a necessary and sufficient condition for (i)–(iii). to hold is that there exists a strictly increasing real-valued function $h(x)$ taking values in [0, 1] such that

$$h\big(M(\neg A|C)\big) = 1 - h\big(M(A|C)\big)$$
$$h\big(M(A \,\&\, B|C)\big) = h\big(M(A|B \,\&\, C)\big)h\big(M(B|C)\big).$$ [18]

Thus a simple rescaling of M gives the finitely-additive probability axioms, and since no other constraint than consistency with the rules of propositional logic is imposed there is no loss of generality in taking the finitely-additive probability axioms themselves as the general solution of (i)–(iii).

It is obvious that there are other rescalings of (i)–(iii) that are not *formally* probability functions, but this does not affect the claim that those axioms represent consistency constraints. It has already been observed that mere rescaling, like the odds rescaling, changes nothing at the fundamental level, for the same reason that arbitrary transformations of the 0 and 1 for 'false' and 'true' will generate different arithmetical representations of the propositional connectives, but otherwise change nothing. More importantly from the present perspective, Cox's result does not extend to an endorsement of countable additivity. We know that the formulas of a first order language can be extended in a reasonably well-behaved way to closure under countable conjunctions and disjunctions, but there is no natural extension of Cox's proof which would exploit that additional structure to yield countable additivity.

[16] Cox was a working physicist and his point of departure was a typical one: to look for *invariant* principles:

> to consider first ... what principles of probable inference will hold however probability is measured. Such principles, if there are any, will play in the theory of probable inference a part like that of Carnot's principle in thermodynamics, which holds for all possible scales of temperature, or like the parts played in mechanics by the equations of Lagrange and Hamilton, which have the same form no matter what system of coordinates is used in the description of motion [4, p. 1].

[17] The differentiability conditions were dispensed with in subsequent derivations by Aczél [1, pp. 320–324] and Paris [19, pp. 24–32].

[18] There has been some controversy about Cox's proof (see, e.g., [13]), and in particular Cox's inference of the associativity of g from his proof that for any A, B, C, D in the background propositional language such that $A \,\&\, B \,\&\, C$ is consistent, $g(g(M(B|C), M(A|B \,\&\, C)), M(D|A \,\&\, B \,\&\, C)) = g(M(B|C), g(M(A|B \,\&\, C), M(D|A \,\&\, B \,\&\, C)))$. But associativity only follows if there are enough propositions whose values approximate any given triple of real numbers in D arbitrarily closely, and to ensure this Paris introduces a further assumption [19, Co 5, p. 24]. However, Cox's differentiability conditions imply that f and g are fixed under changes in values of M for given arguments; in other words we are considering not just actual but also *possible* values of $M(A|C)$, whence associativity follows.

6. Conclusion

There has been much more done in the application of logic to probability, or vice versa if you prefer that way of looking at things, than I have mentioned here. It is less the limitations of space and time that have prevented me from commenting on this work than the desire not to deviate from the heuristic path of seeing the formalism of epistemic probability as a generalisation of the classical-logical concepts of model, consistency and consequence.[19] This explains the time I have devoted to the work of Gaifman and Scott and Krauss, which followed the same path but bounded it more narrowly in terms of generalising the model theory of formal logical languages. This is, I believe, for the reasons given in the foregoing, *too* narrow to give a satisfactory theory, and I have attempted to broaden it by liberalising the class of 'languages' to the usual σ-fields of the familiar mathematical probability theory. To what extent I have been successful it is not of course for me to say.

Acknowledgements

I would like to acknowledge helpful comments by two anonymous referees, and advice and assistance from David Makinson and Jeff Paris.

References

[1] J. Aczél, Functional Equations and their Applications, Academic Press, New York, 1966.
[2] R. Carnap, Logical Foundations of Probability, University of Chicago Press, Chicago, 1950.
[3] G. Coletti, R. Scozzafava, Probabilistic Logic in a Coherent Setting, Kluwer, Dordrecht, 2002.
[4] R.T. Cox, The Algebra of Probable Inference, The Johns Hopkins Press, Baltimore, 1961.
[5] B. de Finetti, La Logique de la Probabilité, in: Actes du Congrès International de Philosophie Scientifique, vol. IV, Hermann, Paris, 1936, pp. 1–9.
[6] B. de Finetti, Foresight, its logical laws, its subjective sources, translated and reprinted in [18], 1937, pp. 53–119.
[7] B. de Finetti, Probability, Induction and Statistics, Wiley, London, 1972.
[8] B. de Finetti, Theory of Probability, vol. 1, Wiley, London, 1974 (English translation of Teoria delle Probabilità, Einaudi, 1970).
[9] H. Gaifman, Concerning measures in first order calculi, Israel Journal of Mathematics 2 (1964) 1–18.
[10] H. Gaifman, M. Snir, Probabilities over rich languages, testing and randomness, Journal of Symbolic Logic 47 (1982) 495–548.
[11] I.J. Good, Probability and the Weighing of Evidence, Charles Griffin, London, 1950.
[12] J.Y. Halpern, An analysis of first-order logics of probability, Artificial Intelligence 46 (1990) 311–350.
[13] J.Y. Halpern, Cox's theorem revisited, Journal of Artificial Intelligence Research 11 (1999) 429–435.
[14] J.B. Kadane, A. O'Hagan, Using finitely additive probability: uniform distributions on the natural numbers, Journal of the American Statistical Association 90 (1995) 626–631.
[15] J.B. Kadane, M.J. Schervish, T. Seidenfeld, Reasoning to a foregone conclusion, Journal of the American Statistical Association 91 (1996) 1228–1235.
[16] H.J. Keisler, J.F. Knight, Barwise infinitary logic and admissible sets, The Bulletin of Symbolic Logic 10 (2004) 4–36.
[17] K. Kelly, The Logic of Reliable Inquiry, Cambridge University Press, Cambridge, 1994.
[18] H. Kyburg, H. Smokler (Eds.), Studies in Subjective Probability, second ed., Wiley, New York, 1980.
[19] J. Paris, The Uncertain Reasoner's Companion, Cambridge University Press, Cambridge, 1994.
[20] L.J. Savage, The Foundations of Statistics, Wiley, New York, 1954.
[21] D. Scott, Logic with denumerably long formulas and finite strings of quantifiers, in: Symposium on the Theory of Models, North-Holland, 1965, pp. 329–341.
[22] D. Scott, P. Krauss, Assigning probabilities to logical formulas, in: J. Hintikka, P. Suppes (Eds.), Aspects of Inductive Logic, North-Holland, Amsterdam, 1966, pp. 219–264.
[23] M.J. Schervish, T. Seidenfeld, J.B. Kadane, State-dependent utilities, Journal of the American Statistical Association 85 (1990) 840–847.
[24] F. Schick, Dutch bookies and money pumps, The Journal of Philosophy 83 (1986) 112–119.
[25] S. Shapiro, Foundations without Foundationalism, Oxford University Press, Oxford, 1991.
[26] R.M. Smullyan, First Order Logic, Dover, New York, 1968.
[27] B. Skyrms, Pragmatics and Empiricism, Yale University Press, New Haven, 1984.
[28] J. Swift, A modest proposal for preventing the children of poor people in Ireland, from being a burden to their parents or country; and for making them beneficial to the public, in: A Tale of a Tub and other Satires, J.M. Dent and Sons, London, 1978, pp. 255–264.
[29] J. Väänänen, Second-order logic and foundations of mathematics, Bulletin of Symbolic Logic 7 (2001).
[30] C. Villegas, On qualitative probability σ-algebras, Annals of Mathematical Statistics 35 (1964) 1787–1796.
[31] J. Williamson, Countable additivity and subjective probability, British Journal for the Philosophy of Science 50 (1999) 401–416.

[19] For example, I have not mentioned, the work of Halpern [12] and others which formalises a type of probabilistic reasoning within a first order framework. This work, with its embedded probability functions, is more a type of quantitative *modal* logic (classical modal logic also of course 'brings down' modal operators into the object language and consequently allows an arbitrary degree of modal embedding).