A Probabilistic Logic Based on the Acceptability of Gambles ⋆

Peter R. Gillett [a,*], Richard B. Scherl [b,1], Glenn Shafer [a]

[a] *Rutgers Business School—Newark and New Brunswick*

[b] *Monmouth University, New Jersey*

Abstract

This article presents a probabilistic logic whose sentences can be interpreted as asserting the acceptability of gambles described in terms of an underlying logic. This probabilistic logic has a concrete syntax and a complete inference procedure, and it handles conditional as well as unconditional probabilities. It synthesizes Nilsson's probabilistic logic and Frisch and Haddawy's anytime inference procedure with Wilson and Moral's logic of gambles.

Two distinct semantics can be used for our probabilistic logic: (1) the measure-theoretic semantics used by the prior logics already mentioned and also by the more expressive logic of Fagin, Halpern, and Meggido and (2) a behavioral semantics. Under the measure-theoretic semantics, sentences of our probabilistic logic are interpreted as assertions about a probability distribution over interpretations of the underlying logic. Under the behavioral semantics, these sentences are interpreted only as asserting the acceptability of gambles, and this suggests different directions for generalization.

Key words:
Probabilistic Logic, Anytime Deduction, Gambles, Measure-theoretic, Behavioral Semantics.

⋆ This research was supported by NSF grant SES-9819116 and has benefitted from conversations with Peter Haddawy, Joe Halpern and Nic Wilson.

∗ Corresponding author.

Email addresses: `gillett@rbsmail.rutgers.edu` (Peter R. Gillett), `rscherl@monmouth.edu` (Richard B. Scherl), `gshafer@andromeda.rutgers.edu` (Glenn Shafer).

[1] Richard Scherl acknowledges additional support from the Knowledge Fusion Center of the Army Research Laboratory under contract number DAAD-03-2-0034 and from ARDA under award number 2004-H900700-000.

1 Introduction

This article presents a probabilistic logic $\mathcal{L}$ whose sentences can be interpreted as asserting the acceptability of gambles. The logic $\mathcal{L}$ has a concrete syntax and a complete inference procedure, and it handles conditional as well as unconditional probabilities. It synthesizes the probabilistic logic of Nils J. Nilsson [10] and the anytime inference procedure of Alan M. Frisch and Peter Haddawy [4] with the logic of gambles of Nic Wilson and Serafín Moral [17].

Nilsson and Frisch and Haddawy build their probabilistic logics, which we designate by $\mathcal{L}_N$ and $\mathcal{L}_{FH}$, respectively, on top of an underlying logic. According to their semantics, which we call the *measure-theoretic semantics* for probabilistic logic, each sentence says something about the probability of a sentence in the underlying logic. Our probabilistic logic, which we designate by $\mathcal{L}$, also has an underlying logic, and can use measure-theoretic semantics. When it does use this semantics, it is a strict generalization of Nilsson's and Frisch and Haddawy's logics: a sentence in $\mathcal{L}_N$ or $\mathcal{L}_{FH}$ translates into $\mathcal{L}$ with no change in meaning. However, $\mathcal{L}$ also contains more complex sentences. Instead of merely saying something about the probability of an individual sentence of the underlying logic, a sentence in $\mathcal{L}$ may say something about the expected value of a gamble whose payoff depends on the truth values of several sentences in the underlying logic.

Moreover, whereas Nilsson only discusses how to reason with models, and Frisch and Haddawy do not demonstrate completeness for their set of inference rules, we give a complete set of inference rules for $\mathcal{L}$. In generalizing from probabilities to expected values, we are following Wilson and Moral, and our demonstration of the completeness of our logic uses the same strategy as a demonstration of the completeness of their logic, which we designate by $\mathcal{L}_{WM}$. We go well beyond their results, however, because we handle conditional as well as unconditional probabilities and we insist on a concrete syntax.

Fagin, Halpern and Meggido [3], have formulated probabilistic logics that use measure-theoretic semantics, and have complete inference procedures; we designate them by $\mathcal{L}_{FHM}$. They consider only the case where the underlying logic is propositional logic, but in this case, their probabilistic logics are more expressive than ours. In the case where our probabilistic logic $\mathcal{L}$ uses measure-theoretic semantics and uses propositional logic as its underlying logic, it can be regarded as a relatively small fragment of one of Fagin, Halpern, and Meggido's logics, but it is still of some interest, because it enables complete inference about relatively elementary probability statements (including those considered by Nilsson and by Frisch and Haddawy) without the greater complexity of Fagin, Halpern, and Meggido's logics. In [6], Halpern and Pucella consider upper probability measures, and in [7] they add reasoning about expectation. There is a large body of other important related work on probabilistic logic that is not directly used in our paper ([2] and [5] include reviews of this literature). Recently, for example, Biazzo, Gilio, Lukasiewicz and Sanfilippo [1] have described an approach to probabilistic logic based on betting schemes, and Lukasiewicz [8] has extended this approach to related nonmonotonic probabilistic logics.

What we find most interesting about $\mathcal{L}$ is an alternative semantics that suggests paths for generalization different from the paths followed by other authors. This alternative semantics, which we call *behavioral semantics*, hews more closely to the notion of acceptability and can be formalized using the concept of lower prevision developed by Peter Walley [15]. In measure-theoretic semantics, an interpretation for a probabilistic logic is a probability distribution over interpretations for the underlying logic. A gamble is acceptable if this probability distribution gives the gamble nonnegative expected value. But this implies that if a gamble is not acceptable, then the opposite gamble (the gamble with the signs of all the payoffs reversed) is acceptable. No such implication is inherent in the notion of a gamble being acceptable, and no such implication is built into our probabilistic logic (the syntax considered in this article does not even provide for negation of acceptability statements). Behavioral semantics avoids these implications and therefore can be extended to martingale trees [11] and probability games [14]; see [13].

1.1 Comparative Summary

Table 1 summarizes how our logic compares with others. As the table indicates, our approach is a synthesis of Nilsson's, Frisch and Haddawy's and Wilson and Moral's, combining the best features of their approaches. We also show for comparison Fagin, Halpern and Meggiddo's more expressive logics.

	$\mathcal{L}_N$	$\mathcal{L}_{FH}$	$\mathcal{L}_{WM}$	$\mathcal{L}_{FHM}$	$\mathcal{L}$
Complete inference procedure	Yes	No	Yes	Yes	Yes
Anytime inference procedure	No	Yes	Yes	Yes	Yes
Elaboration tolerant	No	Yes	No	Yes	Yes
Concrete syntax	No	Yes	No	Yes	Yes
Handles conditional probabilities	No	Yes	No	Yes	Yes
Hyperplane expressive	No	No	Yes	Yes	Yes

Table 1
Comparison of Related Work

When we say that Nilsson's system does not have an anytime inference procedure, we mean that the linear program must be run to completion. In contrast our language, like that of Frisch and Haddawy, is modular, and interim inferences are valid even though they may not have computed the tightest possible bounds at the time computation is stopped.

Another advantage of modularity is elaboration tolerance: additional premises can be introduced and additional questions can be asked without discarding or repeating work already done. Our language, like Frisch and Haddawy's, is elaboration tolerant in this sense. Notice, however, that we have labeled Wilson and Moral's language as elaboration intolerant, even though it seems to have an anytime inference procedure. This is because its inference procedure takes for granted that the sample space has already been set up. When we introduce a new sentence, whether as a premise or a goal, old possibilities may split, according to whether the new sentence is true or false. So no system that takes the sample space for granted is elaboration tolerant.

Our assertion that Wilson and Moral do not have a concrete syntax refers to the fact that they do not specify any particular symbolic representation for their gambles. They specify syntax neither for their probabilistic logic nor for an underlying logic $\mathcal{L}_0$. Nilsson, in contrast, does insist on a concrete syntax for the underlying logic $\mathcal{L}_0$, although he does not specify a syntax for his probabilistic logic.

When we say that Wilson and Moral do not handle conditional probabilities, we mean only that they do not do so explicitly. A bound on a conditional probability can easily be re-expressed as a statement of the type they do handle. When we say that our logic and that of Wilson and Moral are hyperplane expressive, we are referring to the fact that these logics can express an arbitrary linear constraint on a vector of probabilities $\boldsymbol{P}$, which requires $\boldsymbol{P}$ to lie on one side of a hyperplane. Such linear constraints can be much more general than bounds on individual probabilities and conditional probabilities.

Each of these three logics can be seen as a simplification of the logic $\mathcal{L}$ developed in this article. $\mathcal{L}$ is not, however, the most expressive probabilistic logic possible. It can bound the vector of probabilities $\boldsymbol{P}$ by hyperplanes, and hence it can express the statement that $\boldsymbol{P}$ is in a given simplex, but it cannot express more complicated restrictions on $\boldsymbol{P}$. For example, it cannot express statements about the probability p of a single sentence in the underlying logic such as "$0.3 \leq p \leq 0.5$ or $0.5 \leq p \leq 0.7$." Probabilistic logics that can express such statements include those of Fagin, Halpern, and Megiddo [3], which make use of rich inferential machinery including all instances of propositional tautologies, modus ponens, all instances of valid formulae about linear inequalities, and four axioms for probability. Semantically, the absence of Boolean combinations of sentences in our logic is significant; in particular, sentences in our logic assert the (conditional) acceptability of gambles—our logic is not designed to express the assertion that a gamble is (conditionally) unacceptable. Finally, Fagin, Halpern, and Megiddo's logics follow an axiomatic approach, with modus ponens as the single inference rule; in contrast, our logic follows a natural deduction approach with a single axiom and a number of inference rules. In our view, this approach is intuitively simpler.

2 Syntax and Inference for $\mathcal{L}$

We designate our probabilistic logic by $\mathcal{L}$. In this section, we describe $\mathcal{L}$'s syntax and inference procedure. This description involves some informal explanation of $\mathcal{L}$'s semantics. In the next two sections, we formalize the semantics in two different ways and demonstrate the soundness and completeness of the inference procedure with respect to both formalizations.

The sentences of $\mathcal{L}$ have the form

$$\langle (\alpha_1, a_1) \dots (\alpha_n, a_n) \mid \delta \rangle, \tag{1}$$

where n is a nonnegative integer, $\alpha_1, \dots, \alpha_n$ and δ are sentences of an underlying logic $\mathcal{L}_0$, and $a_1, \dots, a_n$ are real numbers. The list $(\alpha_1, a_1) \dots (\alpha_n, a_n)$ represents a gamble, which pays the sum of those a_i for which the corresponding sentence α_i turns out to be true. Sentence (1) means that this gamble is acceptable to an agent when his knowledge relative to the sentences in $\mathcal{L}_0$ consists of knowledge that δ is true. In the next two sections, we make this idea into a formal semantics in two different ways. In §3, we formalize it as the condition that the payoff of the gamble has nonnegative expected value conditional on δ; because the expected value has to be computed relative to some probability distribution, this constitutes a use of what we have already called *measure-theoretic semantics* for probabilistic logic. In §4, we formalize it in terms of our *behavioral semantics*, in which probability distributions are replaced by lower previsions.

If the sentence δ in $\mathcal{L}_0$ is a tautology, then the sentence (1) in $\mathcal{L}$ means that the gamble $(\alpha_1, a_1) \dots (\alpha_n, a_n)$ is acceptable *a priori*. This is the special case of unconditional acceptability considered by Wilson and Moral. But even here we differ from Wilson and Moral by representing gambles in terms of sentences of an underlying logic rather than merely as functions on a sample space.

2.1 The Underlying Logic $\mathcal{L}_0$

We assume that the underlying logic $\mathcal{L}_0$ has propositional symbols p, p', p'', etc., is two-valued with values in $\{$*true*, *false*$\}$, and uses the symbols $\neg$, $\wedge$, $\vee$, $\perp$, and $\top$ in the usual way. In particular, (a) the set of sentences includes each of the propositional symbols, $\perp$, and $\top$, and is closed under $\neg$, $\wedge$ and $\vee$; and (b) an interpretation ω satisfies $\neg\alpha$ if and only if it does not satisfy α, satisfies $\alpha \wedge \beta$ if and only if it satisfies both α and β, satisfies $\alpha \vee \beta$ if and only if it satisfies at least one of α or β, never satisfies $\perp$ and always satisfies $\top$. We use $\Rightarrow_0$ and $\Leftrightarrow_0$, respectively, for derivability and logical equivalence: $\alpha_1 \Rightarrow_0 \alpha_2$ means that α_2 can be derived from α_1 (i.e., $\alpha_1 \vdash_{\mathcal{L}_0} \alpha_2$), and $\alpha_1 \Leftrightarrow_0 \alpha_2$ means that either can be derived from the other (i.e., $\alpha_1 \vdash_{\mathcal{L}_0} \alpha_2$ and $\alpha_2 \vdash_{\mathcal{L}_0} \alpha_1$).

We assume that $\mathcal{L}_0$ has a sound and complete inference procedure, so that $\alpha_1 \Leftrightarrow_0 \alpha_2$ holds whenever the two are semantically equivalent. We assume that $\mathcal{L}_0$'s inference procedure is complete only because we need this assumption in order to show completeness for $\mathcal{L}$. It is not needed in order for $\mathcal{L}$'s inference procedure to be well-defined and sound.

We make no further assumptions about $\mathcal{L}_0$, but reasoning within $\mathcal{L}_0$ is part of reasoning within $\mathcal{L}$ and so details about $\mathcal{L}_0$ are relevant to implementation. Unlike Fagin, Halpern, and Megiddo's logics, however, propositional reasoning takes place only within $\mathcal{L}_0$.

Let us write wff_0 for the set consisting of all sentences of $\mathcal{L}_0$. Given a truth assignment M for $\mathcal{L}_0$, let us designate by ω_M the interpretation it determines—this is a mapping from wff_0 to $\{\mathit{true}, \mathit{false}\}$. And let us write Ω_0 for the set consisting of all such interpretations:

$$\Omega_0 := \{\omega_M \mid M \text{ is a truth assignment for } \mathcal{L}_0\}.$$

We call Ω_0 the *sample space for* $\mathcal{L}_0$. This concept should be contrasted with the notion of the sample space for a finite set of sentences in $\mathcal{L}_0$, used in Nilsson's work. Whereas we might explicitly construct the sample space for a few hundred sentences, there is no reasonable sense in which we can explicitly construct Ω_0. If $\mathcal{L}_0$ is undecidable, explicit construction of Ω_0 is not even theoretically possible. But as a theoretical (rather than a computational) object, Ω_0 will be very useful in our mathematical reasoning about $\mathcal{L}$.

We call any real-valued function on Ω_0 a *variable*, and denote by $\mathcal{X}$ the set of all variables. Given a variable $X \in \mathcal{X}$ and a subset $A \subseteq \Omega_0$, we define a variable X^A by

$$X^A(\omega) := \begin{cases} X(\omega) & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

We call X^A the *restriction* of X to A. We can write $X^A = X \cdot I_A$, where the dot denotes pointwise multiplication and I_A is A's *indicator variable*:

$$I_A(\omega) := \begin{cases} 1 & \text{if } \omega \in A, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

Clearly $(X_1 + X_2)^A = X_1^A + X_2^A$, $(rX)^A = rX^A$, and $(X^A)^B = X^{A \cap B}$.

Given $\alpha \in \mathit{wff}_0$, let $[\alpha]$ be the subset of Ω_0 consisting of those truth assignments that assign α the value *true*: $[\alpha] := \{\omega \in \Omega_0 \mid \omega(\alpha) = \mathit{true}\}$. Then the set $\mathcal{A}_0 := \{[\alpha] \mid \alpha \in \mathit{wff}_0\}$ is a field of subsets of Ω_0; it is isomorphic to $\mathcal{L}_0$'s Lindenbaum-Tarski field. We call a finitely additive probability measure on the field $\mathcal{A}_0$ a *probability distribution* on Ω_0.

We call an ordered pair (α, a), where $\alpha \in \mathit{wff}_0$ and a is a real number, a *ticket.* We call a list $(\alpha_1, a_1) \dots (\alpha_n, a_n)$, where n is a nonnegative integer and the (α_i, a_i) are tickets, a *gamble.* The α_i are the *sentences* of the gamble; the a_i are the *payoffs.* The integer n may be zero; in this case the gamble is an empty list. Notice also that a ticket may occur in a gamble more than once. We write *Gamble* for the set consisting of all gambles. We use meta-variables such as G, G', etc. to designate gambles without specifying their tickets.

Given a gamble $G = (\alpha_1, a_1) \dots (\alpha_n, a_n)$, we define a variable X_G by

$$X_G(\omega) := \sum \{a_i | 1 \leq i \leq n \text{ and } \omega(\alpha_i) = \mathit{true}\} = \sum_{i=1}^{n} a_i \cdot I_{[\alpha_i]}(\omega). \tag{4}$$

We call X_G the *variable representation* for the gamble G. Many different gambles can be represented by the same variable. We call a variable X *simple* if $X = X_G$ for some gamble G. A variable is simple if and only if (1) it takes only a finite number of values, and (2) for each real number r, there exists $\alpha \in \mathit{wff}_0$ such that $\{\omega \mid X(\omega) = r\} = [\alpha]$ (in the language of probability theory, X is measurable with respect to $\mathcal{A}_0$).

Here is our notation for manipulating gambles:

- Given a gamble $G = (\alpha_1, a_1) \dots (\alpha_n, a_n)$ and a sentence β in $\mathcal{L}_0$, we write G^β for the result of conjoining each α_i with β:

$$G^\beta := (\alpha_1 \wedge \beta, a_1) \dots (\alpha_n \wedge \beta, a_n).$$

- Given a gamble $G = (\alpha_1, a_1) \dots (\alpha_n, a_n)$ and a real number r, we write rG for the result of multiplying each a_i by r:

$$rG := (\alpha_1, ra_1) \dots (\alpha_n, ra_n).$$

- Given gambles $G = (\alpha_1, a_1) \dots (\alpha_n, a_n)$ and $G' = (\beta_1, b_1) \dots (\beta_m, b_m)$, we write GG' for the result of concatenating G and G':

$$GG' := (\alpha_1, a_1) \dots (\alpha_n, a_n) (\beta_1, b_1) \dots (\beta_m, b_m).$$

These manipulations affect the variable representation in obvious and straightforward ways: $X_{G^\beta} = X_G^{[\beta]}$ (see equation (2)), $X_{rG} = rX_G$, and $X_{GG'} = X_G + X_{G'}$.

We sometimes want to *append* a ticket to a gamble or *remove* an instance of a ticket from a gamble:

- Appending a ticket to a gamble means adding it at the end, without regard to whether it already occurs in the gamble. For example, the result of appending (β, b) to $(\alpha, a)\,(\beta, b)\,(\gamma, c)$ is $(\alpha, a)\,(\beta, b)\,(\gamma, c)\,(\beta, b)$.
- Removing an instance of (β, b) from $(\alpha, a)\,(\beta, b)\,(\gamma, c)\,(\beta, b)$ can result in either $(\alpha, a)\,(\beta, b)\,(\gamma, c)$ or $(\alpha, a)\,(\gamma, c)\,(\beta, b)$.

Here are ten important ways of changing a gamble we call *elementary moves*:

0	permute the order of the tickets
1	append $(\alpha, 0)$
2	remove an instance of $(\alpha, 0)$
3	append $(\bot, a)$
4	remove an instance of $(\bot, a)$
5	remove an instance of (α, a), and append (α, a_1) and (α, a_2), where $a_1 + a_2 = a$
6	remove an instance of (α, a_1) and an instance of (α, a_2), and append $(\alpha, a_1 + a_2)$
7	remove an instance of (α, a), and append (β, a) and (γ, a), where $\beta \wedge \gamma \Leftrightarrow_0 \bot$ and $\beta \vee \gamma \Leftrightarrow_0 \alpha$
8	remove an instance of (β, a) and an instance of (γ, a), and append (α, a), where $\beta \wedge \gamma \Leftrightarrow_0 \bot$ and $\beta \vee \gamma \Leftrightarrow_0 \alpha$
9	remove an instance of (α, a) and append (β, a), where $\beta \Leftrightarrow_0 \alpha$

We say that two gambles G and G' *equivalent* if we can get from one to the other by elementary moves—i.e., if there is a finite sequence of gambles $G_1, \ldots, G_k$ such that $G_1 = G$, $G_k = G'$, and G_{i+1} can be obtained from G_i by an elementary move, for $i = 1, \ldots, k-1$. This is evidently an equivalence relation. Elementary moves in part are similar to Fagin, Halpern, and Meggido's valid formulae about linear equalities and in part capture intuitions regarding the decomposition of gambles and the nature of probability.

Proposition 2.1 *Gambles G and G' are equivalent if and only if $X_G = X_{G'}$.*

It is easy to see that $X_G = X_{G'}$ when G' is obtained from G by an elementary move, and this implies that $X_G = X_{G'}$ whenever G and G' are equivalent. So our task is to show that $X_G = X_{G'}$ implies the equivalence of G and G'.

To do so, we introduce some additional concepts. We say that two of $\mathcal{L}_0$'s sentences α and β are *disjoint* if $\alpha \wedge \beta \Leftrightarrow_0 \bot$ and that a gamble $(\alpha_1, a_1) \ldots (\alpha_n, a_n)$ is *in standard form* if the following conditions are satisfied:

- The sentences are disjoint: $\alpha_i \wedge \alpha_j \Leftrightarrow_0 \bot$ for $i \neq j$.
- No sentence is absurd: $\alpha_i \not\Leftrightarrow_0 \bot$ for all i.
- No payoff is zero: $a_i \neq 0$ for all i.
- The payoffs are distinct and in increasing order: $a_1 < \cdots < a_n$.

Lemma 2.2 *Any gamble is equivalent to a gamble in standard form.*

Proof: Consider a gamble $G = (\alpha_1, a_1) \dots (\alpha_n, a_n)$. By repeated elementary moves of type 7, we can reduce G to an equivalent gamble G_1 in which every ticket's sentence has the form

$$\beta_1 \wedge \cdots \wedge \beta_n, \tag{5}$$

where for each i, either $\beta_i = \alpha_i$ or else $\beta_i = \neg\alpha_i$. (We ignore the placement of parentheses in the expression (5), but we assume that these parentheses are placed in some canonical way, using if necessary elementary moves of type 9.) Any two sentences in G_1 are either disjoint or equivalent. Using elementary moves of type 6 and 9, we can consolidate the tickets with equivalent sentences, reducing G_1 to an equivalent gamble G_2 whose sentences are disjoint. Using repeated elementary moves of type 8, we can reduce G_2 to an equivalent gamble G_3 whose sentences are still disjoint and whose payoffs are all distinct. Elementary moves of types 0, 2, and 4 will then reduce G_3 to an equivalent gamble in standard form. ■

Lemma 2.3 *If G and G' are in standard form, and $X_G = X_{G'}$, then G and G' are equivalent.*

Proof: As $X_G = X_{G'}$, the two gambles must have the same list of payoffs $a_1, \dots, a_n$. Since G and G' are in standard form, $G = (\alpha_1, a_1) \dots (\alpha_n, a_n)$ and $G' = (\beta_1, a_1) \dots (\beta_n, a_n)$, where $[\alpha_i] = [\beta_i]$ for $i = 1, \dots, n$. So $\alpha_i \Leftrightarrow_0 \beta_i$, and G can be transformed into G' by elementary moves of type 9. ■

We complete the proof of Proposition 2.1 by considering two gambles G_1 and G_2 such that $X_{G_1} = X_{G_2}$ and showing that they are equivalent. By Lemma 2.2, we have gambles G_1' and G_2' that are in standard form and are equivalent to G_1 and G_2 respectively. The equivalence of G_i' and G_i implies that $X_{G_i'} = X_{G_i}$, and hence that $X_{G_1'} = X_{G_2'}$. By Lemma 2.3, G_1' and G_2' are equivalent. Hence G_1 and G_2 are equivalent.

2.3 The Syntax of $\mathcal{L}$

A sentence of $\mathcal{L}$ is any expression of the form (1), where n is a nonnegative integer, δ and $\alpha, \dots, \alpha_n$ are sentences of $\mathcal{L}_0$, and $a_1, \dots, a_n$ are real numbers. Notice that n is allowed to be zero, so that $\langle \mid \delta \rangle$ is a sentence of $\mathcal{L}$.

We write *wff* for the set consisting of all sentences of $\mathcal{L}$. For a sentence $\langle (\alpha_1, a_1) \dots (\alpha_n, a_n) \mid \delta \rangle$ in *wff* we call $(\alpha_1, a_1) \dots (\alpha_n, a_n)$ its *gamble*, and δ its *condition*. We use meta-variables such as S, S', etc. to designate elements of *wff* without specifying their gambles or conditions.

We do not form negations of the sentences in $\mathcal{L}$: when S is a sentence of $\mathcal{L}$, $\neg S$ is *not* a sentence in $\mathcal{L}$. Nor do we form conjunctions or disjunctions.

2.4 Inference in $\mathcal{L}$

We now define an inference relation $\vdash$ for $\mathcal{L}$.

We adopt one axiom schema and five inference rules. The axiom schema applies to any $\alpha, \delta \in \mathit{wff}_0$ while the inference rules apply to any $G, G' \in \mathit{Gamble}$, and any $\delta, \epsilon \in \mathit{wff}_0$.

Acceptability
RATIONALITY $\vdash \langle(\alpha, 1) \mid \delta\rangle$.
SUBSTITUTION $\langle G \mid \delta\rangle \vdash \langle G' \mid \delta\rangle$ if G and G' are equivalent.
COMBINATION $\{\langle G \mid \delta\rangle, \langle G' \mid \delta\rangle\} \vdash \langle GG' \mid \delta\rangle$.
SCALING $\langle G \mid \delta\rangle \vdash \langle rG \mid \delta\rangle$ if $r \geq 0$.
Conditioning
CONTINGENCY $\langle G \mid \delta\rangle \vdash \langle G^\delta \mid \epsilon\rangle$ if $\delta \Rightarrow_0 \epsilon$.
UPDATING $\langle G^\delta \mid \epsilon\rangle \vdash \langle G \mid \delta\rangle$ if $\delta \Rightarrow_0 \epsilon$.

The axiom schema and first three rules capture our notion of the acceptability of gambles, consistent with Wilson and Moral's approach or Walley's sense of desirability as applied to our formal notion of gambles. Thus, a gamble in which we can only win is always acceptable; an acceptable gamble is acceptable no matter how it is written; the combination of acceptable gambles is acceptable, and a multiple or fraction of an acceptable gamble is itself acceptable. The conditioning rules, on the other hand, are exactly what is needed to capture a particular interpretation of conditional probabilities (the "called-off bet" interpretation implicit in Frisch and Haddawy's system). Under this interpretation, conditional probabilities can be defined in terms of unconditional probabilities, and so our logic could be formulated more simply without using conditioning sentences. Whether the conditioning rules appear trivial or deep depends, of course, on one's intuitions regarding the nature of conditional probabilities. Our intention in future work, however, is to study variants of our logic $\mathcal{L}$ in which the acceptability axiom and rules are retained in conjunction with weaker conditioning rules, and conditioning sentences then remain fundamental. In particular, we are interested in studying weaker forms of UPDATING.

Inference proceeds in the usual way. One starts with a set of premises and enlarges it in steps, including at each either an axiom or a sentence whose inference is authorized from sentences already in the set by one of the inference rules. If $\Gamma \subseteq \mathit{wff}$, $S \in \mathit{wff}$, and we can infer S from Γ, then we write $\Gamma \vdash S$.

Implementing the inference rules involves, of course, using the inference procedure of the underlying logic $\mathcal{L}_0$. In order to use CONTINGENCY, for example, we must show that ϵ can be inferred from δ. Inference in $\mathcal{L}_0$ enters even into the use of SUBSTITUTION, since we need to demonstrate equivalence or implication in $\mathcal{L}_0$ in order to prove the equivalence of two gambles.

The following proposition lists some elementary consequences of our inference rules.

Proposition 2.4 *(1)* $\langle G \mid \delta\rangle \vdash \langle G^\delta \mid \delta\rangle$ *and* $\langle G^\delta \mid \delta\rangle \vdash \langle G \mid \delta\rangle$.
(2) $\langle G \mid \delta\rangle \vdash \langle G^\delta \mid \top\rangle$ *and* $\langle G^\delta \mid \top\rangle \vdash \langle G \mid \delta\rangle$.
(3) If $\delta \Leftrightarrow_0 \epsilon$*, then* $\langle G \mid \delta\rangle \vdash \langle G \mid \epsilon\rangle$.

Proof: We obtain the two inferences in Statement 1 by setting ϵ equal to δ in CONTINGENCY and UPDATING, respectively. We similarly obtain Statement 2 by setting ϵ equal to $\top$. To derive Statement 3, we start with $\langle G \mid \delta\rangle$, use CONTINGENCY to get $\langle G^\delta \mid \epsilon\rangle$, use the equivalence of the gambles G^δ and G^ϵ, together with SUBSTITUTION, to get $\langle G^\epsilon \mid \epsilon\rangle$, and then use UPDATING to get $\langle G \mid \epsilon\rangle$. ■

It is noteworthy that the logic $\mathcal{L}$ includes an absurdity—a sentence from which any other sentence can be inferred. This is the sentence $\langle(\top, -1) \mid \top\rangle$. This sentence says that our agent is willing to give away \$1 *a priori*, and the inference rules allow us to infer from this that he will be willing to give away any amount of money under any other state of knowledge δ. If $\Gamma \vdash \langle(\top, -1) \mid \top\rangle$, then we say that Γ is *incoherent*. More generally, if $\Gamma \vdash \langle(\delta, -1) \mid \delta\rangle$, then we say that Γ is *incoherent* in δ.

3 Measure-Theoretic Semantics for $\mathcal{L}$

As we have already explained informally, we can adapt the measure-theoretic semantics for our language $\mathcal{L}$ by using the notion of conditional expected value:

- An *interpretation* of $\mathcal{L}$ is a probability distribution P on the sample space Ω_0.
- An interpretation P *satisfies* a sentence $\langle G \mid \delta\rangle$ if P's expected value for the variable X_G (see equation (4)), conditional on δ, is nonnegative.

This definition of satisfaction is only informal. Our formal definition will resolve the indeterminacy of conditional expected value when the condition δ has probability zero in a way consistent with Frisch and Haddawy: the sentence is satisfied by the interpretation in this case.

In this section, we study the entailment relation for $\mathcal{L}$ based on this measure-theoretic semantics and then show that the inference procedure we described in the preceding section is sound and complete with respect to this relation. We explain our alternative semantics, behavioral semantics in the next section.

3.1 Entailment Under Measure-Theoretic Semantics

Formally, we say that P *satisfies* $\langle(\alpha_1, a_1)\dots(\alpha_n, a_n) \mid \delta\rangle$ if

$$\sum_{i=1}^{n} a_i \cdot P([\alpha_i] \cap [\delta]) \geq 0. \tag{6}$$

This inequality is equivalent to

$$\sum_{i=1}^{n} a_i \cdot \frac{P([\alpha_i] \cap [\delta])}{P([\delta])} \geq 0, \tag{7}$$

provided that we agree to the convention that the ratio $P([\alpha_i] \cap [\delta])/P([\delta])$ is equal to zero (and hence inequality (7) is satisfied) whenever the denominator, $P([\delta])$, is equal to zero. The left-hand side of inequality (7) is the conditional expected value of the variable corresponding to the gamble $(\alpha_1, a_1)\dots(\alpha_n, a_n)$, conditional on the event $[\delta]$. This justifies the informal definition we offered a moment ago: P satisfies $\langle G \mid \delta\rangle$ if the expected value of X_G, conditional on δ, is nonnegative. This treatment of conditional probability is mandated by the called-off bet interpretation of conditional probability discussed earlier.

If we write G for the gamble $(\alpha_1, a_1)\dots(\alpha_n, a_n)$, then we can rewrite the inequality (6) as a condition on the variable representation for G:

$$E_P X_G^{[\delta]} \geq 0, \tag{8}$$

where E_P represents the expected value operator for P.

We write $\models^m$ for the measure-theoretic entailment relation for $\mathcal{L}$: $\Gamma \models^m S$ if and only if P satisfies S whenever P satisfies S' for all $S' \in \Gamma$. As usual, we abbreviate $\Gamma \models^m S$ to $\models^m S$ when Γ is empty; this means that every interpretation P satisfies S.

3.2 Soundness Under Measure-Theoretic Semantics

Now we verify that $\mathcal{L}$'s inference procedure is sound with respect to measure-theoretic semantics: if $\Gamma \vdash G$, then $\Gamma \models^m G$. It suffices to show that the axioms and the inference rules are sound.

RATIONALITY $\vdash \langle(\alpha, 1) \mid \delta\rangle$.

Every interpretation P satisfies $\langle(\alpha, 1) \mid \delta\rangle$, because inequality (6) reduces to $P([\alpha] \cap [\delta]) \geq 0$ for this sentence, and a probability is always nonnegative.

SUBSTITUTION $\langle G \mid \delta\rangle \vdash \langle G' \mid \delta\rangle$ if G and G' are equivalent.

Soundness follows from the fact that inequality (6) can be put in the form (8) and the fact that equivalent gambles have the same variable representation (Proposition 2.1).

COMBINATION $\{\langle G \mid \delta\rangle, \langle G' \mid \delta\rangle\} \vdash \langle GG' \mid \delta\rangle$.

Soundness follows from inequality (8) and the relation $X_{GG'}^{[\delta]} = X_G^{[\delta]} + X_{G'}^{[\delta]}$.

SCALING $\langle G \mid \delta\rangle \vdash \langle rG \mid \delta\rangle$ if $r \geq 0$.

Soundness follows from inequality (8) and the relation $X_{rG}^{[\delta]} = rX_G^{[\delta]}$.

CONTINGENCY $\langle G \mid \delta\rangle \vdash \langle G^\delta \mid \epsilon\rangle$ if $\delta \Rightarrow_0 \epsilon$.

Here we use inequality (8) and the calculation $X_{G^\delta}^{[\epsilon]} = (X_G^{[\epsilon]})^{[\delta]} = X_G^{[\epsilon]\cap[\delta]}$; when $\delta \Rightarrow_0 \epsilon$, $[\epsilon] \cap [\delta] = [\delta]$.

UPDATING $\langle G^\delta \mid \epsilon\rangle \vdash \langle G \mid \delta\rangle$ if $\delta \Rightarrow_0 \epsilon$.

Soundness follows by the same argument as for CONTINGENCY.

3.3 Completeness Under Measure-Theoretic Semantics

We first establish several results that we use later in our demonstration:

Lemma 3.1 *If $X_G^{[\delta]} = X_{G'}^{[\delta']}$, then $\langle G \mid \delta\rangle \vdash \langle G' \mid \delta'\rangle$.*

Proof: We have $\langle G \mid \delta\rangle \vdash \langle G^\delta \mid \delta \vee \delta'\rangle$ by CONTINGENCY. Our hypothesis $X_G^{[\delta]} = X_{G'}^{[\delta']}$ implies $X_{G^\delta} = X_{G'^{\delta'}}$. So by Proposition 2.1, G^δ and $G'^{\delta'}$ are equivalent, and therefore $\langle G^\delta \mid \delta \vee \delta'\rangle \vdash \langle G'^{\delta'} \mid \delta \vee \delta'\rangle$ by SUBSTITUTION. Finally, $\langle G'^{\delta'} \mid \delta \vee \delta'\rangle \vdash \langle G' \mid \delta'\rangle$ by UPDATING. ■

Lemma 3.2 *If $r > 0$ and $X_{G'}^{[\delta']} = rX_G^{[\delta]}$, then $\langle G \mid \delta\rangle \vdash \langle G' \mid \delta'\rangle$.*

Proof: We have $\langle G \mid \delta\rangle \vdash \langle rG \mid \delta\rangle$ by SCALING. And because our hypothesis can be written in the form $X_{G'}^{[\delta']} = X_{rG}^{[\delta]}$, we have $\langle rG \mid \delta\rangle \vdash \langle G' \mid \delta'\rangle$ by Lemma 3.1. ■

Lemma 3.3 *If $X_G^{[\delta]} = X_{G_1}^{[\delta_1]} + X_{G_2}^{[\delta_2]}$, then $\{\langle G_1 \mid \delta_1\rangle, \langle G_2 \mid \delta_2\rangle\} \vdash \langle G \mid \delta\rangle$.*

Proof: Because $X_{G_i}^{[\delta_i]} = X_{G_i{}^{\delta_i}} = X_{G_i{}^{\delta_i}}^{[\delta_1 \vee \delta_2]} (i = 1, 2)$, we can infer $\langle G_1^{\delta_1} \mid \delta_1 \vee \delta_2\rangle$ and $\langle G_2^{\delta_2} \mid \delta_1 \vee \delta_2\rangle$ by Lemma 3.1. We can then infer $\langle G_1^{\delta_1} G_2^{\delta_2} \mid \delta_1 \vee \delta_2\rangle$ by COMBINATION. Because $X_{G_1{}^{\delta_1} G_1{}^{\delta_1}}^{[\delta_1 \vee \delta_2]} = X_{G_1{}^{\delta_1}}^{[\delta_1 \vee \delta_2]} + X_{G_2{}^{\delta_2}}^{[\delta_1 \vee \delta_2]} = X_{G_1}^{[\delta_1]} + X_{G_2}^{[\delta_2]} = X_G^{[\delta]}$, we can then infer $\langle G \mid \delta\rangle$ by Lemma 3.1. ■

In the style used by Wilson [16] for an extended version of Wilson and Moral's logic, we sketch the outline of a proof that $\mathcal{L}$'s inference procedure is complete under our measure-theoretic semantics: if Γ is finite and $\Gamma \models^m S$, then $\Gamma \vdash S$.

It is convenient to consider the variable representation of gambles, and to rely on the fact that our gambles are finite: each contains only a finite number of tickets. A finite Γ therefore involves only a finite set of sentences in wff_0 that use only a finite set of propositional symbols $\{p_i\}_{i=1}^k$. We write wff_* for the set of sentences generated by $\{p_i\}_{i=1}^k$. Given a truth assignment M_* for $\{p_i\}_{i=1}^k$, we designate by ω_{M_*} the interpretation it determines—a mapping from wff_* to $\{\mathit{true}, \mathit{false}\}$. We write $\Omega_* := \{\omega_{M_*} \mid M_* \text{ is a truth assignment for } \{p_i\}_{i=1}^k\}$ and $\mathcal{A}_* := \{[\alpha]_* \mid \alpha \in \mathit{wff}_*\}$. Clearly, there are only a finite number of different interpretations determined in this way; i.e, Ω_* and $\mathcal{A}_*$ are finite. For each $\omega \in \Omega_0$ there is a $\omega_* \in \Omega_*$ that agrees with ω for each member of wff_*, and vice versa. For each finitely additive probability measure P on $\mathcal{A}_0$ there is a finitely additive probability measure P_* on $\mathcal{A}_*$ that agrees with it on $\mathcal{A}_*$, and vice versa. Each P_* is defined by its values on the individual $\omega \in \Omega_*$. Note that for any $S \in \Gamma$, P satisfies S if and only if P_* satisfies S. Hence, the P_* form a new set of restricted models for $\mathcal{L}$.

A sentence $\langle G \mid \delta \rangle \in \Gamma$ has the variable representation $X_G^{[\delta]}$, a real-valued function on Ω_0. For each such variable $X_G^{[\delta]}$ we can identify a variable $X_{*G}^{[\delta]}$ which is a real-valued function on Ω_* such that $X_G^{[\delta]}(\omega) = X_{*G}^{[\delta]}(\omega_*)$. We write S_* for the variable on Ω_* generated in this way by a sentence $S \in \Gamma$ when its gamble and condition are unspecified. Though we proved them in their more general form, Lemmas 3.1 to 3.3 apply also to the restricted variables on Ω_*.

We write $\mathcal{I}_*$ for the set of indicator variables $\mathcal{I}_* := \{I_{\{w\}} \mid w \in \Omega_*\}$. For each $\omega \in \Omega_*$ there is a sentence $\alpha_\omega \in \mathit{wff}_*$ such that the indicator variable on Ω_* $I_{\{\omega\}} = X_{G_\omega} = X_{G_\omega}^{[\top]}$, where $G_\omega := (\alpha_\omega, 1)$; by Rationality, $\vdash \langle(\alpha_\omega, 1) \mid \top\rangle$.

The set of variables on Ω_* forms a linear space, which we denote by $\mathcal{X}$. A subset of $\mathcal{X}$ that is closed under addition and multiplication by non-negative scalars is called a convex cone. Given $\Gamma \subseteq \mathcal{X}$, we write $C(\Gamma)$ for the smallest convex cone containing Γ:

$$C(\Gamma) := \{r_1 X_1 + \ldots + r_n X_n | n \geq 1, 0 \leq r_i \in \mathbb{R}, X_i \in \Gamma\}.$$

We call a convex cone Δ finite if $\Delta = C(\Gamma)$ for a finite Γ. We define the inner product of two variables X and Y by $XY := \sum_{\omega \in \Omega_*} X(\omega)Y(\omega)$, and we define the dual cone Γ^+ of a subset $\Gamma \subseteq \mathcal{X}$ by:

$$\Gamma^+ := \{X \in \mathcal{X} \mid XY \geq 0 \text{ for all } Y \in \Gamma\}.$$

We make use of a number of properties of finite cones (see Nering [9]):

Theorem 3.4 (Reflexivity of finite convex cones) *If Δ is a finite convex cone, then it is reflexive; i.e. $(\Delta^+)^+ = \Delta$.*

This theorem follows from a number of fundamental results on finite cones and dual cones [9]. We also use the following property of dual cones:

Theorem 3.5 *For any $\Gamma_1 \subseteq \mathcal{X}$ and $\Gamma_2 \subseteq \mathcal{X}$, if $\Gamma_1 \subseteq \Gamma_2$ then $\Gamma_1^+ \supseteq \Gamma_2^+$*

In order to apply these results to establish the finite completeness of our logic $\mathcal{L}$, we first define, for a set of sentences Γ:

the restricted variable representation of Γ, $\Gamma_* := \{X_{*G}^{[\delta]} \mid \langle G \mid \delta \rangle \in \Gamma\}$
the set of syntactic consequences of Γ, $\mathit{Con}_{\vdash} := \{X_{*G}^{[\delta]} \mid \Gamma \vdash \langle G \mid \delta \rangle\}$;
the set of semantic consequences of Γ, $\mathit{Con}_{\models} := \{X_{*G}^{[\delta]} \mid \Gamma \models^m \langle G \mid \delta \rangle\}$; and
the set of non-negative functions, $\mathcal{R} := \{R \in \mathcal{L} \mid R(\omega) \geq 0 \text{ for all } \omega \in \Omega_*\}$.

Proposition 3.6 *For a finite set of sentences Γ:*

(1) $\mathit{Con}_{\models}(\Gamma) = (\Gamma_*^+ \cap \mathcal{R})^+$
(2) $\Gamma_*^+ \cap \mathcal{R} = \mathit{Con}_{\vdash}(\Gamma)^+ \cap \mathcal{R}$
(3) $\mathit{Con}_{\vdash}(\Gamma) = C(\Gamma_* \cup \mathcal{I}_*)$
(4) $(C(\Gamma_* \cup \mathcal{I}_*))^+ \cap \mathcal{R} = (C(\Gamma_* \cup \mathcal{I}_*))^+$
(5) $\mathit{Con}_{\models}(\Gamma) = (\mathit{Con}_{\vdash}(\Gamma)^+)^+$
(6) $\mathit{Con}_{\models}(\Gamma) = \mathit{Con}_{\vdash}(\Gamma)$

Proof:

(1) $\mathit{Con}_{\models}(\Gamma) = (\Gamma_*^+ \cap \mathcal{R})^+$

$$\begin{aligned} \mathit{Con}_{\models}(\Gamma) &= \{X_{*G}^{[\delta]} \mid P \text{ satisfies } \langle G \mid \delta \rangle \text{ whenever } P \text{ satisfies } Y \\ &\quad \text{for every } Y \in \Gamma\} \\ &= \{X_{*G}^{[\delta]} \mid P_* \text{ satisfies } \langle G \mid \delta \rangle \text{ whenever } P_* \text{ satisfies } Y \\ &\quad \text{for every } Y \in \Gamma\} \\ &= \{X \in \mathcal{X} \mid TX \geq 0 \text{ for every } T \in \mathcal{R} \text{ such that } TY \geq 0 \\ &\quad \text{for every } Y \in \Gamma_*\} \\ &= (\Gamma_*^+ \cap \mathcal{R})^+, \text{ by the definition of dual cones.} \end{aligned}$$

(2) $\Gamma_*^+ \cap \mathcal{R} = \mathit{Con}_{\vdash}(\Gamma)^+ \cap \mathcal{R}$

$$\begin{aligned} \Gamma_*^+ \cap \mathcal{R} &\subseteq \{T \mid TY \geq 0 \text{ for every } Y \in \Gamma_*, \text{ where } T \text{ is} \\ &\quad \text{a non-negative multiple of a probability measure}\} \\ &\subseteq \{T \mid TY \geq 0 \text{ for every } Y \in \mathit{Con}_{\vdash}(\Gamma), \text{ where } T \text{ is} \\ &\quad \text{a non-negative multiple of a probability measure}\} \\ &\quad \text{since by Soundness if } Y \in \mathit{Con}_{\vdash}(\Gamma) \text{ then } Y \in \mathit{Con}_{\models}(\Gamma) \\ &\subseteq \mathit{Con}_{\vdash}(\Gamma)^+ \cap \mathcal{R}. \end{aligned}$$

$\Gamma_*^+ \supseteq \mathit{Con}_{\vdash}(\Gamma)^+$ by Theorem 3.5 as $\Gamma_* \subseteq \mathit{Con}_{\vdash}(\Gamma)$ by definition, so $\Gamma_*^+ \cap \mathcal{R} \supseteq \mathit{Con}_{\vdash}(\Gamma)^+ \cap \mathcal{R}$.
Hence, $\Gamma_*^+ \cap \mathcal{R} = \mathit{Con}_{\vdash}(\Gamma)^+ \cap \mathcal{R}$.

(3) $\mathit{Con}_{\vdash}(\Gamma) = C(\Gamma_* \cup \mathcal{I}_*)$

$\mathit{Con}_{\vdash}(\Gamma) \supseteq C(\Gamma_* \cup \mathcal{I}_*)$ from the definitions, the observation that the members of $\mathcal{I}_*$ are variable representations of theorems, and Lemmas 3.2 and 3.3.

$\mathit{Con}_{\vdash}(\Gamma) \subseteq C(\Gamma_* \cup \mathcal{I}_*)$ from the definitions, the observation that $X_{*(\alpha,1)}^{[\delta]} = I_{[\alpha\cap\delta]}$, the fact that by Proposition 2.1 equivalent gambles are represented by the same variable, and the observation that CONTINGENCY and UPDATING do not change the variable representation of sentences as shown by the calculations used above in establishing Soundness.

(4) $(C(\Gamma_* \cup \mathcal{I}_*))^+ \cap \mathcal{R} = (C(\Gamma_* \cup \mathcal{I}_*))^+$

$(C(\Gamma_* \cup \mathcal{I}_*))^+ = \{T \mid TX \geq 0 \text{ for every } X \in C(\Gamma_* \cup \mathcal{I}_*)\}$.

For every $\omega \in \Omega_*$, $I_{\{\omega\}} \in C(\Gamma_* \cup \mathcal{I}_*)$ and so for every $T \in (C(\Gamma \cup \mathcal{I}))^+$, $T(\omega) \geq 0$ for every $\omega \in \Omega_*$; i.e., for every $T \in (C(\Gamma \cup \mathcal{I}))^+$, $T \in \mathcal{R}$.

Thus, $(C(\Gamma \cup \mathcal{I}))^+ \cap \mathcal{R} = (C(\Gamma \cup \mathcal{I}))^+$.

(5) $\mathit{Con}_{\models}(\Gamma) = (\mathit{Con}_{\vdash}(\Gamma)^+)^+$

By applying (1), (2), (3), (4), and (3) in turn.

(6) $\mathit{Con}_{\models}(\Gamma) = \mathit{Con}_{\vdash}(\Gamma)$

By (3), $\mathit{Con}_{\vdash}(\Gamma)$ is a finite convex cone, and so by Theorem 3.4 $\mathit{Con}_{\vdash}(\Gamma) = (\mathit{Con}_{\vdash}(\Gamma)^+)^+$; hence, by (5) $\mathit{Con}_{\models}(\Gamma) = \mathit{Con}_{\vdash}(\Gamma)$

■

To complete our demonstration of finite completeness, suppose $\Gamma \models^m S$, where

$$\Gamma = \{\langle G_1 \mid \delta_1\rangle, \ldots, \langle G_k \mid \delta_k\rangle\} \quad \text{and} \quad S = \langle G \mid \delta\rangle.$$

Then $X_{*G}^{[\delta]} \in \mathit{Con}_{\models}(\Gamma)$, and so by Proposition 3.6 (6) $X_{*G}^{[\delta]} \in \mathit{Con}_{\vdash}(\Gamma)$; i.e., there is a sentence $\langle G' \mid \delta'\rangle$ such that $\Gamma \vdash \langle G' \mid \delta'\rangle$ and $X_{*G'}^{[\delta']} = X_{*G}^{[\delta]}$. By Lemma 3.1, $\langle G' \mid \delta'\rangle \vdash \langle G \mid \delta\rangle$. Consequently, $\Gamma \vdash \langle G \mid \delta\rangle$; i.e., $\Gamma \vdash S$.

4 Behavioral Semantics for $\mathcal{L}$

In our behavioral semantics for $\mathcal{L}$, which we study in this section, we replace the concept of a probability distribution with a concept of lower prevision adapted from Walley [15]. The context for this section is the syntax and inference procedure for $\mathcal{L}$ that we developed in §2. In particular, an underlying logic $\mathcal{L}_0$ is in place, and the sample space Ω_0 is defined. We leave aside, however, the semantics developed in §3 to study a different semantics for the same syntax and inference procedure.

Suppose Ω is a nonempty set and $\mathcal{A}$ is a field of subsets of Ω. Write $\mathcal{X}$ for the linear space consisting of all real-valued functions on Ω that are measurable with respect to $\mathcal{A}$ and take only finitely many values. We call a real-valued function $\underline{P}$ on $\mathcal{X}$ a *lower prevision* for Ω if it satisfies these three conditions:

(1) $\underline{P}(X) \geq \inf\{X(\omega) \mid \omega \in \Omega\}$ for all $X \in \mathcal{X}$.
(2) $\underline{P}(X_1 + X_2) \geq \underline{P}(X_1) + \underline{P}(X_2)$ for all $X_1, X_2 \in \mathcal{X}$.
(3) $\underline{P}(rX) = r\underline{P}(X)$ for all $r \geq 0$ and $X \in \mathcal{X}$.

We call a lower prevision satisfying condition 2 with equality for all $X_1, X_2 \in \mathcal{X}$ a *linear prevision*.

The following proposition gives some insight into the concepts of lower prevision and linear prevision by relating them to more familiar concepts.

Proposition 4.1 *(1) A real-valued function $\underline{P}$ on $\mathcal{X}$ is a linear prevision if and only if there is a finitely additive probability measure P on $(\Omega, \mathcal{A})$ such that*

$$\underline{P}(X) = E_P X \text{ for all } X \in \mathcal{X}.$$

(Here E_P is the expected value operator for P.)

(2) A real-valued function $\underline{P}$ on $\mathcal{X}$ is a lower prevision if and only if it is the lower envelope of the expected value operators for a set of finitely additive probability measures—i.e., if and only if there is a set Λ of finitely additive probability measures on $(\Omega, \mathcal{A})$ such that

$$\underline{P}(X) = \inf_{P \in \Lambda} E_P X \text{ for all } X \in \mathcal{X}.$$

Statement 1 is proven in §3.2 of Walley [15] and Statement 2 in §3.3. Notice, however, that our terminology is not quite the same as Walley's. He calls any real-valued function on any set of bounded real-valued functions on Ω a lower prevision, and he relates the three conditions that we have used as the definition of lower prevision to a concept that he calls coherence. For an explanation of our disagreement with Walley regarding coherence, see [13].

Our motivation for considering lower previsions does not derive from their relation to probability measures. On the contrary, we regard lower previsions as more fundamental than probability measures, because they emerge more directly from the idea of the acceptability of gambles. For the moment, let us follow Wilson and Moral by thinking of the elements of $\mathcal{X}$ as gambles, and let us write $\mathcal{C}$ for the subset of $\mathcal{X}$ consisting of the gambles we consider acceptable. What should $\mathcal{C}$ look like? According to the intuition that underlies both our inference procedure for $\mathcal{L}$ and our definition of lower prevision, $\mathcal{C}$ should satisfy three conditions (see also Walley [15]):

(1) If $X_1, \dots, X_n \in \mathcal{C}$, and $r_1, \dots, r_n$ are positive real numbers, then $r_1 X_1 + \dots + r_n X_n \in \mathcal{C}$.
(2) If $\inf_{\omega \in \Omega} X(\omega) \geq 0$, then $X \in \mathcal{C}$.
(3) $-1 \notin \mathcal{C}$.

But it is easy to verify that if $\mathcal{C}$ verifies these conditions, and we set

$$\underline{P}(X) = \sup\{a \mid X - a \in \mathcal{C}\} \tag{9}$$

for all $X \in \mathcal{X}$, then $\underline{P}$ qualifies as a lower prevision, and

$$\underline{P}(X) \geq 0 \text{ if and only if } X \in \mathcal{C}.$$

(Accepting $X - a$ is the same as paying a for X, and so (9) is the most one will pay for X. Thus $\underline{P}(X) \geq 0$ if and only if one will pay 0 for X.)

Conversely, if we start with a lower prevision $\underline{P}$, and set $\mathcal{C} := \{X \mid \underline{P}(X) \geq 0\}$, then $\mathcal{C}$ will satisfy our three conditions. So we could define our behavioral semantics directly in terms of $\mathcal{C}$ rather than $\underline{P}$. This would make behavioral semantics much more transparent. Using lower previsions has the advantage, however, of emphasising the similarities with measure-theoretic semantics.

4.2 Entailment Under Behavioral Semantics

Here are the definitions of interpretation and satisfaction in our behavioral semantics for $\mathcal{L}$:

- An *interpretation* is a lower prevision on Ω_0.
- An interpretation $\underline{P}$ *satisfies* a sentence $\langle G \mid \delta \rangle$ if

$$\underline{P}(X_G^{[\delta]}) \geq 0. \tag{10}$$

This is quite parallel to measure-theoretic semantics, where an interpretation, a probability distribution P on Ω_0, satisfies $\langle G \mid \delta \rangle$ if $E_P(X_G^{[\delta]}) \geq 0$. Because the expected value operators for probability distributions are a special kind of lower prevision (namely, linear previsions), we may say that behavioral semantics generalizes measure-theoretic semantics. We write $\models^b$ for the entailment relation for logic $\mathcal{L}$ determined by this new definition of satisfaction.

4.3 Soundness Under Behavioral Semantics

The demonstration that $\mathcal{L}$'s axioms and inference rules are sound with respect to behavioral semantics proceeds just like the demonstration with respect to measure-theoretic semantics (§3.2).

4.4 Completeness Under Behavioral Semantics

The completeness of $\mathcal{L}$ under behavioral semantics follows easily from its completeness under measure-theoretic semantics.

Suppose, indeed, that $\Gamma \models^b S$, where

$$\Gamma = \{\langle G_1 \mid \delta_1\rangle, \ldots, \langle G_k \mid \delta_k\rangle\} \quad \text{and} \quad S = \langle G \mid \delta\rangle.$$

This means that if $\underline{P}$ is a lower prevision and $\underline{P}(X_{G_i}^{[\delta_i]}) \geq 0$ for $i = 1, \ldots, k$, then $\underline{P}(X_G^{[\delta]}) \geq 0$. Because a linear prevision is a lower prevision, this means in particular that if $E_P X_{G_i}^{[\delta_i]} \geq 0$ for $i = 1, \ldots, k$, then $E_P X_G^{[\delta]} \geq 0$. In other words, $\Gamma \models^m S$, where $\models^m$ is the entailment relation under measure-theoretic semantics. So we obtain $\Gamma \vdash S$ from completeness under measure-theoretic semantics.

5 Expressibility and Proof in $\mathcal{L}$

A sentence in $\mathcal{L}_N$ expresses a bound on the probability of a sentence in the underlying logic, while a sentence in $\mathcal{L}_{FH}$ expresses bounds on the conditional probability of such a sentence (in this case, a sentence in propositional logic). Any bound on the probability or conditional probability of a sentence can be expressed as the condition that a particular gamble has a nonnegative expected value, and therefore both $\mathcal{L}_N$ and $\mathcal{L}_{FH}$ can be regarded as fragments of our more expressive logic $\mathcal{L}$. There is however, more to say, especially in the case of $\mathcal{L}_{FH}$, which has formal inference rules. Just how do sentences in $\mathcal{L}_N$ or $\mathcal{L}_{FH}$ translate into sentences in $\mathcal{L}$, and how are $\mathcal{L}_{FH}$'s inference rules related to $\mathcal{L}$'s?

Nilsson, in $\mathcal{L}_N$, began with sentences of the form $\mathbb{P}(\alpha) = a$ and inferred sentences of the more general forms $a \leq \mathbb{P}(\alpha)$ and $\mathbb{P}(\alpha) \leq b$. In order to translate these sentences into $\mathcal{L}$, we can be guided by a fact about the common semantics: the probability of an event A under an interpretation P is the same as the expected value under P of the indicator variable I_A. This produces the translations shown in Table 2.

Sentence in $\mathcal{L}_N$	Equivalent condition on P	Sentence in $\mathcal{L}$
$a \leq \mathbb{P}(\alpha)$	$E_P(I_{[\alpha]} - a) \geq 0$	$\langle(\alpha, 1)\,(\top, -a) \mid \top\rangle$
$\mathbb{P}(\alpha) \leq b$	$E_P(b - I_{[\alpha]}) \geq 0$	$\langle(\top, b)\,(\alpha, -1) \mid \top\rangle$

Table 2
Translating from Nilsson's logic $\mathcal{L}_N$ to $\mathcal{L}$.

In Frisch and Haddawy's logic $\mathcal{L}_{FH}$, a sentence simultaneously expresses a lower and an upper bound on a conditional probability:

$$\mathbb{P}(\alpha \mid \delta) \in [a, b]. \tag{11}$$

How should we translate this sentence into $\mathcal{L}$? The most natural approach might be to add the condition δ to the translations of $a \leq \mathbb{P}(\alpha)$ and $\mathbb{P}(\alpha) \leq b$ in Table 2. This produces two sentences:

$$\langle(\alpha, 1)\,(\top, -a) \mid \delta\rangle \text{ and } \langle(\top, b)\,(\alpha, -1) \mid \delta\rangle. \tag{12}$$

Another approach is to translate the sentence (11) directly into a condition on an interpretation P:

$$a \leq \frac{P([\alpha] \cap [\delta])}{P([\delta])} \leq b. \tag{13}$$

Under the convention that the ratio is zero when its denominator is zero, condition (13) is equivalent to the two conditions

$$P([\alpha] \cap [\delta]) - a \cdot P([\delta]) \geq 0 \text{ and } b \cdot P([\delta]) - P([\alpha] \cap [\delta]) \geq 0,$$

and these two conditions are naturally expressed in $\mathcal{L}$ by the two sentences

$$\langle(\alpha \wedge \delta, 1)\,(\delta, -a) \mid \top\rangle \text{ and } \langle(\delta, b)\,(\alpha \wedge \delta, -1) \mid \top\rangle \tag{14}$$

Both (12) and (14) are correct; the sentences $\langle(\alpha, 1)\,(\top, -a) \mid \delta\rangle$ and $\langle(\alpha \wedge \delta, 1)\,(\delta, -a) \mid \top\rangle$ are equivalent to each other, and the sentences $\langle(\top, b)\,(\alpha, -1) \mid \delta\rangle$ and $\langle(\delta, b)\,(\alpha \wedge \delta, -1) \mid \top\rangle$ are equivalent to each other. Because the sentences (11) in $\mathcal{L}_{FH}$ have the same interpretation as the sentences (14) in $\mathcal{L}$ (both mean that the condition (13) holds under the convention that the ratio is zero when the denominator is zero), this translation is in fact a translation of $\mathcal{L}_{FH}$ into our probabilistic logic. So we may say that we have extended Frisch and Haddawy's logic and added a sound and complete inference procedure. Together with the soundness of their inference rules, this implies that their inference rules are consequences of ours.

There is one complication, deriving from the fact one of Frisch and Haddawy's is represented by two of our sentences. Because of their representation, Frisch and Haddawy introduce the following inference rule:

$$\frac{\begin{array}{l}\mathbb{P}(\alpha \mid \delta) \in [x, y] \\ \mathbb{P}(\alpha \mid \delta) \in [u, v]\end{array}}{\mathbb{P}(\alpha \mid \delta) \in [max(x, u), min(y, v)]}$$

This rule can be seen as an embodiment of the anytime character of their logic $\mathcal{L}_{FH}$; if we apply this rule whenever it can be applied, we always know the tightest bounds on the probability of α given δ that are justified by our computation so far. Because we express the lower and upper bounds separately, we have no need for such an inference rule, but in an implementation we could, of course, track the largest a for which $\langle(\alpha\wedge\delta, 1)\,(\delta, -a) \mid \top\rangle$ is in our database and the smallest b for which $\langle(\delta, b)\,(\alpha \wedge \delta, -1) \mid \top\rangle$ is in it.

Consider now, for example, the axioms for probability used in the logic of Fagin, Halpern and Meggido [3]:

P1 $\mu(X) \geq 0$ for all $X \in \mathcal{X}$
P2 $\mu(S) = 1$
P3″ $\mu(X) = \mu(X \cap Y) + \mu(X \cap \overline{Y})$

We may write **P3″** equivalently as: if $\mathbb{P}(\alpha \wedge \beta) = a$ and $\mathbb{P}(\alpha \wedge \overline{\beta}) = b$, then $\mathbb{P}(\alpha) = a{+}b$. Then, translating each equality into two inequalities and applying (12), we obtain the following sequent:

(1a)	$\langle(\alpha \wedge \beta, 1)\,(\top, -a) \mid \top\rangle$
(1b)	$\langle(\top, a)\,(\alpha \wedge \beta, -1) \mid \top\rangle$
(2a)	$\langle(\alpha \wedge \overline{\beta}, 1)\,(\top, -b) \mid \top\rangle$
(2b)	$\langle(\top, b)\,(\alpha \wedge \overline{\beta}, -1) \mid \top\rangle$
(3a)	$\langle(\alpha, 1)\,(\top, -(a + b)) \mid \top\rangle$
(3b)	$\langle(\top, a + b)\,(\alpha, -1) \mid \top\rangle$

The derivations of the conclusions are straight-forward. Here is a derivation of (3a) in $\mathcal{L}$:

(i)	$\langle(\alpha \wedge \beta, 1)\,(\top, -a) \mid \top\rangle$	1a
(ii)	$\langle(\alpha \wedge \overline{\beta}, 1)\,(\top, -b) \mid \top\rangle$	2a
(iii)	$\langle(\alpha \wedge \beta, 1)\,(\top, -a)\,(\alpha \wedge \overline{\beta}, 1)\,(\top, -b) \mid \top\rangle$	(i), (ii) Combination
(iv)	$\langle(\alpha \wedge \beta, 1)\,(\alpha \wedge \overline{\beta}, 1)\,(\top, -(a + b)) \mid \top\rangle$	(iii) Substitution type 6
(v)	$\langle((\alpha \wedge \beta) \vee (\alpha \wedge \overline{\beta}), 1)\,(\top, -(a + b)) \mid \top\rangle$	(iv) Substitution type 8
(vi)	$\langle((\alpha, 1)\,(\top, -(a + b) \mid \top\rangle$	(v) Substitution type 9

The derivation of (3b) in $\mathcal{L}$ follows similar lines:

(i)	$\langle(\top, a)\,(\alpha \wedge \beta, -1) \mid \top\rangle$	1b
(ii)	$\langle(\top, b)\,(\alpha \wedge \overline{\beta}, -1) \mid \top\rangle$	2b
(iii)	$\langle(\top, a)\,(\alpha \wedge \beta, -1)\,(\top, b)\,(\alpha \wedge \overline{\beta}, -1) \mid \top\rangle$	(i), (ii) Combination
(iv)	$\langle(\top, a + b)\,(\alpha \wedge \beta, -1)\,(\alpha \wedge \overline{\beta}, -1, \mid \top\rangle$	(iii) Substitution type 6
(v)	$\langle(\top, a + b)\,((\alpha \wedge \beta) \vee (\alpha \wedge \overline{\beta}), -1) \mid \top\rangle$	(iv) Substitution type 8
(vi)	$\langle(\top, a + b)\,((\alpha, -1) \mid \top\rangle$	(v) Substitution type 9

Although these proofs are trivial, it is of interest to note that they use only our

Acceptability rules, and in a sense reveal why **P3**$''$ is a consequence of our notion of acceptability of gambles. Since our logic $\mathcal{L}$ is complete, we know that all Frisch and Haddawy's inference rules can be derived in $\mathcal{L}$. Nevertheless, it may provide some insight into the operation of our logic to show *how* they may be derived. For example, Frisch and Haddawy's Rule (vii) says:

$$\frac{\begin{array}{l}\mathbb{P}(\beta \mid \delta) \in [x, y] \\ \mathbb{P}(\alpha \mid \beta \wedge \delta) \in [u, v]\end{array}}{\mathbb{P}(\alpha \wedge \beta \mid \delta) \in [x \cdot u, y \cdot v]}$$

In $\mathcal{L}$, this can be written as the sequent:

$$\frac{\begin{array}{ll}(1a) & \langle(\beta, 1)\,(\top, -x) \mid \delta\rangle \\ (1b) & \langle(\top, y)\,(\beta, -1) \mid \delta\rangle \\ (2a) & \langle(\alpha, 1)\,(\top, -u) \mid \beta \wedge \delta\rangle \\ (2b) & \langle(\top, v)\,(\alpha, -1) \mid \beta \wedge \delta\rangle\end{array}}{\begin{array}{ll}(3a) & \langle(\alpha \wedge \beta, 1)\,(\top, -(x \cdot u)) \mid \delta\rangle \\ (3b) & \langle(\top, (y \cdot v))\,(\alpha \wedge \beta, -1) \mid \delta\rangle\end{array}}$$

The first half of this sequent, (3a), may be derived in $\mathcal{L}$ as follows:

(i)	$\langle(\beta, 1)\,(\top, -x) \mid \delta\rangle$	1a
(ii)	$\langle(\alpha, 1)\,(\top, -u) \mid \beta \wedge \delta\rangle$	2a
(iii)	$\langle(\alpha \wedge \beta \wedge \delta, 1)\,(\top \wedge \beta \wedge \delta, -u) \mid \delta\rangle$	(ii) CONTINGENCY $\beta \wedge \delta \Rightarrow_0 \delta$
(iv)	$\langle(\alpha \wedge \beta, 1)\,(\top \wedge \beta, -u) \mid \delta\rangle$	(iii) UPDATING $\delta \Rightarrow_0 \delta$
(v)	$\langle(\alpha \wedge \beta, 1)\,(\beta, -u) \mid \delta\rangle$	(iv) SUBSTITUTION type 9
(vi)	$\langle(\beta, u)\,(\top, -(x \cdot u) \mid \delta\rangle$	(i) SCALING $u \geq 0$
(vii)	$\langle(\alpha \wedge \beta, 1)\,(\beta, -u)\,(\beta, u)\,(\top, -(x \cdot u) \mid \delta\rangle$	(v), (vi) COMBINATION
(viii)	$\langle(\alpha \wedge \beta, 1)\,(\beta, 0)\,(\top, -(x \cdot u) \mid \delta\rangle$	(vii) SUBSTITUTION type 6
(ix)	$\langle(\alpha \wedge \beta, 1)\,(\top, -(x \cdot u) \mid \delta\rangle$	(viii) SUBSTITUTION type 2

The structure of the derivation of the final part of the sequent, (3b) follows the same lines and in the interests of space is left as an exercise for the reader. In the derivations shown here, each of our inference rules has played some part, though we have chosen examples where the use of UPDATING is trivial (cf. line iv in the derivation above, where we rely only on $\delta \Rightarrow_0 \delta$); in this article we do not explore the significance and power of this particular form of UPDATING or the consequences of adopting weaker versions instead. However, our axiom of RATIONALITY has not been needed. We leave it to the reader to confirm that it is required, for example, to prove Frisch and Haddawy's Rule (v).

6 Summary and Conclusions

Why stop with a language that is only hyperplane expressive? Why not further expand the language so that it can say anything one pleases about the probabilities of sentences in the underlying logic? There are two obvious ways to answer this question:

- If we are really only interested in the probabilities of individual sentences, as Nilsson and Frisch and Haddawy appeared to be, then there is no reason to generalize further.
- We might feel that we want more than bounds on individual probabilities, but that hyperplane expressiveness is enough. Most practical work with probabilities is directed towards decision-making, and we might argue that decisions depend only on the acceptability of gambles.

We might also challenge the ontological role of probabilities. Is a probability something with a reality of its own, or is it only a way of expressing our attitudes? If it is only a way of expressing our attitudes, and if the attitudes in question come down to the willingness to accept gambles, then we step outside what is meaningful when we go beyond hyperplane expressiveness. This is the view taken by Walley.

Walley's view throws into question, of course, the presumption that a probabilistic logic should use probability measures as interpretations. If the reality to which we are referring has only acceptable gambles, then perhaps these, not probabilities, should be our interpretations. Perhaps also some of our inference rules should be reconsidered. In particular, the conditioning rule UPDATING can be called into question, for it does not have a clear direct justification in terms of the acceptability of gambles.

Our work in this area is motivated by our interest in moving beyond standard probability measures as a semantics for probabilistic logic. In addition to relaxing inference rules such as UPDATING to investigate alternative formulations for conditional probability, we are also interested in shifting away from the static framework of a sample space. This would move probabilistic logic in the direction of temporal and causal logic, and would make contact with our earlier work on basing logic on the concept of an event space [12]. See [13].

References

[1] Veronica Biazzo, Angelo Gilio, Thomas Lukasiewicz, and Giuseppe Sanfilippo. Probabilistic logic under coherence, model-theoretic probabilistic logic, and

default reasoning in system P. *Journal of Applied Non-Classical Logics*, 12(2):189–213, 2002.

[2] Vijay Chandru and John N. Hooker. *Optimization Methods for Logical Inference*. John Wiley & Sons, Inc., New York, 1999.

[3] Ronald Fagin, Joseph Y. Halpern, and Nimrod Megiddo. A logic for reasoning about probabilities. *Information and Control*, 87:78–128, 1990.

[4] Alan Frisch and Peter Haddawy. Anytime deduction for probabilistic logic. *Artificial Intelligence*, 69:93–122, 1994.

[5] Joseph Y. Halpern. *Reasoning about Uncertainty*. The MIT Press, Cambridge, Massachusetts, 2003.

[6] Joseph Y. Halpern and Riccardo Pucella. A logic for reasoning about upper probabilities. In *Proceedings of the 17th Conference on Uncertainty in Artificial Intelligence*, pages 203–210, 2001.

[7] Joseph Y. Halpern and Riccardo Pucella. Reasoning about expectation. In *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence*, pages 207–215, 2002.

[8] Thomas Lukasiewicz. Weak nonmonotonic probabilistic logics. *Artificial Intelligence*, 168(1):119–161, 2005.

[9] D.E. Nering. *Linear Algebra and Matrix Theory*. John Wiley and Sons, 1970.

[10] Nils J. Nilsson. Probabilistic logic. *Artificial Intelligence*, 28:71–87, 1986.

[11] Glenn Shafer. *The Art of Causal Conjecture*. The MIT Press, Cambridge, Massachusetts, 1996.

[12] Glenn Shafer, Peter R. Gillett, and Richard Scherl. The logic of events. *Annals of Mathematics and Artificial Intelligence*, 28:315–389, 2000.

[13] Glenn Shafer, Peter R. Gillett, and Richard B. Scherl. A new understanding of subjective probability and its generalization to lower and upper prevision. *International Journal of Approximate Reasoning*, 33:1–49, 2003.

[14] Glenn Shafer and Vladimir Vovk. *Probability and Finance: It's only a Game!* John Wiley & Sons,Inc., New York, 2001.

[15] Peter Walley. *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall, London, 1991.

[16] Nic Wilson. A logic of extended probability. In *Proceedings of the 1st International Symposium on Imprecise Probabilities and Their Applications, Ghent, Belgium, 29 June - 2 July*, 1999.

[17] Nic Wilson and Serafín Moral. A logical view of probability. In *Proceedings of the 11th European Conference on Artificial Intelligence*. John Wiley & Sons, Ltd, 1994.