Paraconsistent Probabilities: Consistency, Contradictions and Bayes' Theorem

Juliana Bueno-Soler [1,†] and Walter Carnielli [2,*,†]

[1]   Faculty of Technology, State University of Campinas –UNICAMP, Campinas 13484-332, Brazil; juliana@ft.unicamp.br
[2]   Centre for Logic, Epistemology and the History of Science and Department of Philosophy, State University of Campinas—UNICAMP, Campinas 13083-859, Brazil
[*]   Correspondence: walter.carnielli@cle.unicamp.br; Tel.: +55-19-3521-6518
[†]   These authors contributed equally to this work.

Abstract: This paper represents the first steps towards constructing a paraconsistent theory of probability based on the Logics of Formal Inconsistency (LFIs). We show that LFIs encode very naturally an extension of the notion of probability able to express sophisticated probabilistic reasoning under contradictions employing appropriate notions of conditional probability and paraconsistent updating, via a version of Bayes' theorem for conditionalization. We argue that the dissimilarity between the notions of inconsistency and contradiction, one of the pillars of LFIs, plays a central role in our extended notion of probability. Some critical historical and conceptual points about probability theory are also reviewed.

Keywords: paraconsistency; probability; contradiction; consistency; logics of formal inconsistency

---

Quod facile est in re, id probabile est in mente.

—Gottfried Wilhelm Leibniz [1]

1. Consistency versus Non-Contradictoriness

As is widely recognized, paraconsistency is the investigation of logic systems endowed with a negation ¬, such that not every contradiction of the form p and ¬p entails everything. In other terms, a paraconsistent logic does not suffer from trivialism, in the sense that a contradiction does not necessarily explode or trivialize the deductive machinery of the system. In strict terms, even an irrelevant contradiction in traditional logic obliges a reasoner that follows such a logic to derive anything from a pair of contradictory statements $\{\alpha, \neg\alpha\}$, as a result of the so-called Principle of Explosion (PEx):

$$\alpha, \neg\alpha, \vdash \beta, \text{ for arbitrary } \beta.$$

On the other hand, a paraconsistent logician, by using a more cautious way of reasoning, is free of the burden of PEx and could pause to investigate the causes for the contradiction, instead of injudiciously deriving unwanted consequences from it.

Common sense, however, recognizes that some contradictions may be intolerable, and those would destroy the very act of reasoning (that is, lead to trivialization). This amounts to admitting that not all contradictions are equivalent. The Logics of Formal Inconsistency (LFIs), a family of paraconsistent logics designed to express the notion of consistency (and inconsistency, as well) within the object language by employing a connective ∘ (reading ∘α as "α is consistent") realizes such an intuition.

As defended in [2], LFIs can be regarded as theories of logical consequence, epistemic in character, that tell us how to make sensible inferences in the presence of contradictions. From the purely mathematical viewpoint, LFIs are subsystems of classical logic, albeit they extend classical logic in the sense that classicality may be recovered in the presence of consistency: contradictions involving consistent sentences will lead to explosive triviality. Consistency in the LFIs is not regarded as synonymous with freedom from contradiction (as happens with the traditional notion of the consistency of a theory T, where consistency means that there is no sentence α such that $T \vdash \alpha$ and $T \vdash \neg\alpha$ where $\vdash$ is a specified consequence relation in the language of T). The usual notion of consistency, totally depending on negation, is perhaps sufficient for certain mathematical purposes, but not for the whole enterprise of reasoning, as argued in [3]. In the LFIs, however, the notions of consistency and non-contradiction are not coincident, nor are the notions of inconsistency and contradiction the same. For more details, conceptual motivations and the main results about LFIs, the reader is referred to [4,5].

The distinguishing feature of the LFIs is that the principle of explosion is not valid in general, although this law is not abolished, but restricted to consistent sentences. Therefore, a contradictory theory may be non-trivial unless the contradiction refers to something consistent.

Such features of the LFIs are condensed in the following law, which is referred to as the "principle of Gentle Explosion" (PGE):

$$\circ\alpha, \alpha, \neg\alpha \vdash \beta, \text{ for every } \beta, \text{ although } \alpha, \neg\alpha \nvdash \beta, \text{ for some } \beta.$$

More than three decades ago, Williams J. N. [6] pleaded that it is a mistake to suppose that inconsistency is the same as contradiction. The LFIs fully formalize this intuition, and starting from this perspective, it is possible to build a number of logical systems with different assumptions that not only encode classical reasoning, but also (at the price of adding new principles) converge to classical logic. We have chosen a particular logic endowed with adequate principles to deal with our paraconsistent probability measures, without obfuscating the fact that several other logics would give rise to specific (weaker or stronger) notions of paraconsistent probabilistic measures.

2. Ci, a Logic of Formal Inconsistency

The system **Ci** is a member of the hierarchy of the LFIs with some features that make it reasonably close to classical logic; it is appropriate, in this way, to define a generalized notion of probability strong enough to enjoy useful properties. Consider the following stock of propositional axioms and rules:

Definition 1. *Let Σ be a propositional signature closed under the unary connectives $\{\circ, \neg\}$ and under the binary connectives $\{\rightarrow, \wedge, \vee\}$. The logic **Ci** (over Σ) is defined by the following Hilbert calculus:*

Axiom Schemas

Ax1 $\alpha \rightarrow (\beta \rightarrow \alpha)$
Ax2 $(\alpha \rightarrow \beta) \rightarrow ((\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow (\alpha \rightarrow \gamma))$
Ax3 $\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta))$
Ax4 $(\alpha \wedge \beta) \rightarrow \alpha$
Ax5 $(\alpha \wedge \beta) \rightarrow \beta$
Ax6 $\alpha \rightarrow (\alpha \vee \beta)$
Ax7 $\beta \rightarrow (\alpha \vee \beta)$
Ax8 $(\alpha \rightarrow \gamma) \rightarrow ((\beta \rightarrow \gamma) \rightarrow ((\alpha \vee \beta) \rightarrow \gamma))$
Ax9 $\alpha \vee (\alpha \rightarrow \beta)$
Ax10 $\alpha \vee \neg\alpha$
Ax11 $\circ\alpha \rightarrow (\alpha \rightarrow (\neg\alpha \rightarrow \beta))$
Ax12 $\neg\neg\alpha \rightarrow \alpha$
Ax13 $\alpha \rightarrow \neg\neg\alpha$
Ax14 $\neg \circ \alpha \rightarrow (\alpha \wedge \neg\alpha)$

Inference Rule

Modus Ponens (MP): $\frac{\alpha, \ \alpha \rightarrow \beta}{\beta}$

As investigated in detail in [7], **Ci** can be extended to the first-order logic **QCi** (over a convenient extension of Σ) by adding appropriate axioms and rules.

It is worth noting that axioms **Ax1** to **Ax9** plus Modus Ponens **MP** define a Hilbert calculus for positive propositional classical logic (see [4]), and therefore, all of the laws concerning positive logic (as distribution of $\wedge$ over $\vee$, etc.) are valid.

It is instructive to show, as an example, the useful properties of the distribution of conjunction over disjunction that holds as good as in classical logic (not a surprise, since positive classical logic is incorporated into our paraconsistent logic). However, as the validity of such laws may raise some doubts, we provide a quick proof of them. The symbol $\vdash_{\mathbf{Ci}}$ stands for the derivability relation (when there is no danger of confusion, we shall drop the index in $\vdash_{\mathbf{Ci}}$ in order to simplify notation) of **Ci**, and $\alpha \equiv \beta$ means $\alpha \vdash_{\mathbf{Ci}} \beta$ and $\beta \vdash_{\mathbf{Ci}} \alpha$:

Theorem 1. *Distributing conjunctions and disjunctions:*

  1. $\alpha \wedge (\beta \vee \gamma) \equiv (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$
  2. $\alpha \vee (\beta \wedge \gamma) \equiv (\alpha \vee \beta) \wedge (\alpha \vee \gamma)$

Proof. We just prove the first item (the second item is analogous). From right to left, axioms **Ax4** and **Ax5**, plus **Ax8**, easily show the intended implication. From left to right, just notice that:

$$\alpha, \beta \vdash_{\mathbf{Ci}} (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$$

and:

$$\alpha, \gamma \vdash_{\mathbf{Ci}} (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$$

hence, by **Ax8**,

$$\alpha, \beta \vee \gamma \vdash_{\mathbf{Ci}} (\alpha \wedge \beta) \vee (\alpha \wedge \gamma)$$

therefore:

$$\alpha \wedge (\beta \vee \gamma) \vdash_{\mathbf{Ci}} (\alpha \wedge \beta) \vee (\alpha \wedge \gamma).$$

$\square$

As proven in [4], the logic **Ci** cannot be semantically characterized by finite matrices, but it can be characterized in terms of valuations over $\{0, 1\}$ (also called bivaluations):

Definition 2. *Let $\mathcal{L}$ be the collection of sentences of* **Ci**. *A function $v : \mathcal{L} \to \{0, 1\}$ is a valuation for* **Ci** *if it satisfies the following clauses:*

(Bival1) $v(\alpha \wedge \beta) = 1 \iff v(\alpha) = 1$ *and* $v(\beta) = 1$
(Bival2) $v(\alpha \vee \beta) = 1 \iff v(\alpha) = 1$ *or* $v(\beta) = 1$
(Bival3) $v(\alpha \to \beta) = 1 \iff v(\alpha) = 0$ *or* $v(\beta) = 1$
(Bival4) $v(\alpha) = 0 \implies v(\neg\alpha) = 1$
(Bival5) $v(\circ\alpha) = 1 \implies v(\alpha) = 0$ *or* $v(\neg\alpha) = 0$
(Bival6) $v(\neg\neg\alpha) = 1 \iff v(\alpha) = 1$
(Bival7) $v(\neg \circ \alpha) = 1 \implies v(\alpha) = 1$ *and* $v(\neg\alpha) = 1$

The semantical consequence relation w.r.t. bivaluations for **Ci** is defined as expected: $\Gamma \vDash_{\mathbf{Ci}} \varphi$ iff, for every valuation v for **Ci**, if $v(\gamma) = 1$ for every $\gamma \in \Gamma$, then $v(\varphi) = 1$.

Theorem 2. *Completeness theorem:*
For every set of sentences $\Gamma \cup \{\varphi\} \subseteq \mathcal{L}$, $\Gamma \vdash_{\mathbf{Ci}} \varphi$ if and only if $\Gamma \vDash_{\mathbf{Ci}} \varphi$.

Proof. See [4] for a proof. $\square$

As shown in [4], a strong (classical) negation can be defined in **Ci** as $\sim_\beta \alpha = \alpha \to \perp_\beta$, where $\perp_\beta = (\beta \wedge (\neg\beta \wedge \circ\beta))$ is a bottom formula (that is: $\perp_\beta \vdash_{\mathbf{Ci}} \psi$ for every ψ) for any sentence β. In order to simplify matters, a privileged β will be chosen, and the subscript will be omitted in $\perp_\beta$ and $\sim_\beta$ from now on.

Theorem 3. *Properties of strong negation:*
The strong negation $\sim$ satisfies the following properties in **Ci***:*

1. $\vdash_{\mathbf{Ci}} \sim\alpha \to (\alpha \to \psi)$ *for every α and ψ*
2. $\vdash_{\mathbf{Ci}} \alpha \vee \sim\alpha$
3. $\vdash_{\mathbf{Ci}} \alpha \to \sim\sim\alpha$ *and* $\vdash \sim\sim\alpha \to \alpha$
4. *If* $(\Gamma \vdash_{\mathbf{Ci}} \alpha \to \gamma)$ *and* $(\Delta \vdash_{\mathbf{Ci}} \sim\alpha \to \gamma)$, *then* $(\Gamma, \Delta \vdash_{\mathbf{Ci}} \gamma)$
5. $\vdash_{\mathbf{Ci}} (\alpha \to \beta) \to (\sim\beta \to \sim\alpha)$ *and so* $\alpha \to \beta \vdash_{\mathbf{Ci}} \sim\beta \to \sim\alpha$
6. $\vdash_{\mathbf{Ci}} (\sim\alpha \to \sim\beta) \to (\beta \to \alpha)$, *and so,* $\sim\alpha \to \sim\beta \vdash \beta \to \alpha$
7. $\vdash (\alpha \to \sim\beta) \to (\beta \to \sim\alpha)$
8. $\vdash (\varphi \to (\alpha \to \beta)) \to (\varphi \to (\sim\beta \to \sim\alpha))$
9. $\vdash (\varphi \to (\sim\alpha \to \sim\beta)) \to (\varphi \to (\beta \to \alpha))$
10. $\vdash (\varphi \to (\alpha \to \sim\beta)) \to (\varphi \to (\beta \to \sim\alpha))$
11. $\sim\alpha \to \beta \vdash \sim\beta \to \alpha$
12. $\vdash \sim(\alpha \to \beta) \to (\alpha \wedge \sim\beta)$
13. $\vdash \perp \to \alpha$

Proof. Detailed proofs can be found in [4]. □

Additionally, several metatheorems, such as the deduction metatheorem, can be proven in **Ci**.

Now, by defining 'α is inconsistent' as $\bullet\alpha := \neg \circ \alpha$, axioms **Ax12** and **Ax13** (which permit to add and eliminate double negations) convey the meaning that 'α is not inconsistent if and only if it is inconsistent'. Of course, by the very definition and the same axioms on double negations, it also holds 'α is not consistent if and only if it is inconsistent'.

Some other relevant results in **Ci** hold as follows.

Theorem 4. *Properties of consistency:*

1. $\circ\alpha \vdash \neg(\alpha \wedge \neg\alpha)$, *but the converse does not hold*
2. $\circ\alpha \vdash \neg(\neg\alpha \wedge \alpha)$, *but the converse does not hold*
3. $\vdash \circ \circ \alpha$
4. $\vdash \circ \bullet \alpha$
5. $\vdash \neg \bullet \circ\alpha$
6. $\vdash \neg \bullet \bullet\alpha$
7. $\vdash \circ\alpha \vee \alpha$
8. $\vdash \circ\alpha \vee \neg\alpha$
9. $\vdash \circ\alpha \vee \alpha \wedge \neg\alpha$
10. $\alpha \vdash \alpha \wedge (\beta \vee \neg\beta)$
11. $\alpha \wedge (\beta \vee \neg\beta) \vdash \alpha$

Proof. Again, detailed proofs can be found in [4]. □

Theorem 5. *The following are bottom particles in* **Ci***:*

1. $\alpha \wedge \neg\alpha \wedge \circ\alpha \vdash \beta$, *for any β*
2. $\circ\alpha \wedge \bullet\alpha \vdash \beta$, *for any β*
3. $\circ\alpha \wedge \neg \circ \alpha \vdash \beta$, *for any β*
4. $\bullet\alpha \wedge \neg \bullet \alpha \vdash \beta$, *for any β*

Proof. We refer the reader once again to [4] for the proofs. □

3. Consistency, Inconsistency and Paraconsistent Probability

As hinted in the previous section (see [2] for details), the formal notion of consistency here considered does not necessarily depend on negation. Indeed, the logical machinery of the LFIs shows that consistency may be conceived of as a primitive concept, whose meaning can be thought of as "conclusively established as true (or false)" by extra-logical means, depending on the subject matter. Consistency, in this sense, is a notion independent of model theoretical and proof-theoretical methods and is more close to the idea of regularity, or something contrary to change (an elaboration on this view is offered in [8]).

However, consistency is also connected to complying with the laws of traditional probability, as put by F.Ramsey (see [9]), who argues that degrees of belief should satisfy the probability axioms and defends that this is connected to a notion of consistency or coherence. In this way, the notion of consistency (at least for degrees of belief) can be regarded as the satisfaction of the probability axioms.

This paper aims to investigate deeper relationships between logic and probability, emphasizing a new way to define a paraconsistent theory of probability. A previous approach has been developed in [10], where variations of paraconsistent Bayesianism based on a four-valued paraconsistent logic are discussed. A still earlier attempt has been briefly sketched in [11]. An entirely different view is taken in [12], where probabilistic semantics is given for a couple of many-valued paraconsistent logics. The connections between non-classical logics and probability are not circumscribed to paraconsistent logic; see, e.g., [13] for the case of infinite-valued logic and in particular [14] for a broader and more philosophical perspective.

Probability functions are usually defined for a σ-algebra of subsets of a given universe set Ω, but it is also natural to define probability functions directly for sentences in the object language. We will refer to them, respectively, as probability on sets versus probability on sentences (see the discussion in Section 5).

Although these two approaches are equivalent in classical logic in view of the representation theorems for Boolean algebras, this is not so for probability based on other logics, since the algebraic kinship may be lost for non-classical logics or be much less immediate. Furthermore, in algebraic terms, probability functions in set-theoretical settings are required to satisfy countable additivity, but since propositional language is compact, for probability on sentences, it suffices to require finite additivity.

Our first definition of paraconsistent probability will be directly concerned with probability on sentences, with the primary intention to emphasize the role of a new, more cautious logic, behind the probabilistic reasoning, and the effects of logical machinery on a corresponding version of Bayes' rule. In Section 5, however, a notion of paraconsistent probability spaces (concerning probability on sets) will be presented and discussed.

Definition 3. *A probability function for the language $\mathcal{L}$ of a logic **L**, or a **L**-probability function, is a function $P : \mathcal{L} \mapsto \mathbb{R}$ satisfying the following conditions, where $\vdash_L$ stands for the syntactic derivability relation of **L**:*

   1. Non-negativity: $0 \leq P(\varphi) \leq 1$ for all $\varphi \in \mathcal{L}$

   2. Tautologicity: If $\vdash_L \varphi$, then $P(\varphi) = 1$

   3. Anti-tautologicity: If $\varphi \vdash_L$, then $P(\varphi) = 0$

   4. Comparison: If $\psi \vdash_L \varphi$, then $P(\psi) \leq P(\varphi)$

   5. Finite additivity: $P(\varphi \vee \psi) = P(\varphi) + P(\psi) - P(\varphi \wedge \psi)$

It should be remarked that the same meta-axioms, taking appropriately $\vdash_L$ for L the classical, intuitionistic or paraconsistent derivability relation (in the present case, $\vdash_{\mathbf{Ci}}$), define, respectively, the classical, the intuitionistic and the probabilistic probability measures. The intuitionistic case is

treated in [15]. This is clear evidence that the concept of probability can be regarded as entirely logic dependent and that the choice of the underlying logic is a matter of interest and convenience. However, once a choice is made, the consequences are radically different, as we intend to illustrate below.

Two events α and β are said to be independent if $P(\alpha \wedge \beta) = P(\alpha) \cdot P(\beta)$. Two events can be independent concerning one probability measure and dependent concerning another.

Some immediate consequences of such axioms are the following:

Theorem 6. *Regularity of* **L***-probability measures.*

　1. *If φ is a bottom particle in* **L***, then $P(\varphi) = 0$.*
　2. *If φ and ψ are logically equivalent in* **L** *in the sense that $\varphi \vdash \psi$ and $\psi \vdash \varphi$, then $P(\psi) = P(\varphi)$.*

Proof. (1) Immediate, in view of axioms anti-tautologicity and comparison above; (2) Immediate, in view of axiom comparison above.　□

As a consequence of Theorem 6, $P(\alpha \wedge \neg \alpha \wedge \circ \alpha) = 0$, $P(\circ \alpha \wedge \bullet \alpha) = 0$, $P(\circ \alpha \wedge \neg \circ \alpha) = 0$ and $P(\bullet \alpha \wedge \neg \bullet \alpha) = 0$, for any probability function P.

Two sentences α and β are said to be logically incompatible if $\alpha, \beta \vdash \varphi$, for any φ (or equivalently, if $\alpha \wedge \beta$ act as a bottom particle). Some simple calculation rules follow:

Theorem 7. *Calculation rules of* **Ci***-probability measures.*
Let P be a **Ci***-probability measure; then:*

　1. $P(\alpha \vee \beta) = P(\alpha) + P(\beta)$, *if α and β are logically incompatible.*
　2. $P(\circ \alpha) = 2 - (P(\alpha) + P(\neg \alpha))$
　3. $P(\alpha \wedge \neg \alpha) = P(\alpha) + P(\neg \alpha) - 1$
　4. $P(\sim \alpha) = 1 - P(\alpha)$
　5. $P(\neg \circ \alpha) = 1 - P(\circ \alpha)$

Proof. Only Items (1) and (2) will be proven (the rest is routine): (1): Since α and β are logically incompatible, $\alpha \wedge \beta$ act as a bottom particle, and the result is immediate by Theorem 6 and finite additivity; (2): Use finite additivity in the sentences $\circ \alpha \vee (\alpha \wedge \neg \alpha)$ and $\circ \alpha \wedge (\alpha \wedge \neg \alpha)$.　□

Probabilities are sometimes seen as generalized truth values. In this way, in the classical case, the so-called probabilistic semantics may be thought of as extending the bivaluations $v : L \mapsto \{0, 1\}$ of classical propositional logic with the probability functions ranging on the real unit interval $[0, 1]$. The other way round, bivaluations can be regarded as degenerate probability functions; in this sense, classical logic is to be regarded as a special case of probability logic. We show below that an analogous property holds for **Ci** and for the above defined notion of paraconsistent probability.

Probabilistic semantics aim to interpret logic systems (viz., logic entailment) with no appeal to truth conditions. In this way, it differs from standard truth-valued semantics: probabilistic semantics is much more general, notwithstanding both being equivalent in the classical case. What happens is that each semantics expresses logical truth and logical consequence in its own way, and we show in this section that, quite surprisingly, the equivalence between such two distinct semantics also holds for the paraconsistent probability theory based on **Ci**.

Define $\Vdash_P$ as a probabilistic semantic relation whose meaning is $\Gamma \Vdash_P \varphi$ if and only if for every probability function P, if $P(\psi) = 1$ for every $\psi \in \Gamma$, then $P(\varphi) = 1$. It can be shown that **Ci** is (strongly) sound and complete with respect to such probabilistic semantics:

Theorem 8. *Completeness of* **Ci** *with respect to probabilistic semantics:*
$\Gamma \vdash \varphi$ if and only if $\Gamma \Vdash_P \varphi$

Proof. The left-to-right direction follows directly from the axioms of probability, namely, tautologicity and comparison, plus the compactness property of **Ci** proofs. For the right-to-left direction, notice

that if $\Gamma \Vdash_P \varphi$, then in particular, this holds for the probability functions P_2, such that $P_2 : L \mapsto \{0, 1\}$. It suffices, then, by appealing to Theorem 2, to show that the mappings P_2 satisfy all of the conditions for bivaluations of Definition 2:

(Bival1) $P_2(\alpha \wedge \beta) = 1 \iff P_2(\alpha) = 1$ and $P_2(\beta) = 1$

If $P_2(\alpha \wedge \beta) = 1$, since $P_2(\alpha \wedge \beta) \leq P_2(\alpha)$ and $P_2(\alpha \wedge \beta) \leq P_2(\beta)$ by comparison (Definition 3) then $P_2(\alpha) = 1$ and $P_2(\beta) = 1$.

On the other hand, if $P_2(\alpha) = 1$ and $P_2(\beta) = 1$, since again by comparison $P_2(\alpha) \leq P_2(\alpha \vee \beta)$, then $P_2(\alpha \vee \beta) = 1$, and the result follows by finite additivity, i.e., $P(\alpha \wedge \beta) = P(\alpha) + P(\beta) - P(\alpha \vee \beta) = 1$.

(Bival2) $P_2(\alpha \vee \beta) = 1 \iff P_2(\alpha) = 1$ or $P_2(\beta) = 1$

Analogous to the previous item, mutatis mutandis.

(Bival3) $P_2(\alpha \rightarrow \beta) = 1 \iff P_2(\alpha) = 0$ or $P_2(\beta) = 1$

By finite additivity $P_2(\alpha \vee (\alpha \rightarrow \beta)) + P_2(\alpha \wedge (\alpha \rightarrow \beta)) = P_2(\alpha) + P_2(\alpha \rightarrow \beta)$. Since $\vdash \alpha \vee (\alpha \rightarrow \beta)$, tautologicity implies $P_2(\alpha \vee (\alpha \rightarrow \beta)) = 1$, and thus, as $P_2(\alpha \wedge (\alpha \rightarrow \beta)) \leq P_2(\beta)$, $P_2(\beta) \geq P_2(\alpha) + P_2(\alpha \rightarrow \beta) - 1$.

Now, suppose $P_2(\alpha \rightarrow \beta) = 1$; then, either $P_2(\alpha) = 0$ or, if $P_2(\alpha) = 1$, then $P_2(\beta) = 1$.

Reciprocally, suppose that either $P_2(\alpha) = 0$ or $P_2(\beta) = 1$. Since

$P_2(\alpha \vee (\alpha \rightarrow \beta)) + P_2(\alpha \wedge (\alpha \rightarrow \beta)) = 1 + P_2(\alpha \wedge (\alpha \rightarrow \beta)) = P_2(\alpha) + P_2(\alpha \rightarrow \beta)$, if $P_2(\alpha) = 0$, then $P_2(\alpha \rightarrow \beta) \neq 0$, hence $P_2(\alpha \rightarrow \beta) = 1$.

If $P_2(\beta) = 1$, since $\beta \vdash (\alpha \rightarrow \beta)$, it follows immediately $P_2(\alpha \rightarrow \beta) = 1$.

(Bival4) $P_2(\alpha) = 0 \implies P_2(\neg\alpha) = 1$

Suppose that $P_2(\alpha) = 0$. If $P_2(\neg\alpha) = 0$, then by Theorem 7, Item (3), it follows that $P_2(\alpha \wedge \neg\alpha) = -1$, absurd.

(Bival5) $P_2(\circ\alpha) = 1 \implies P_2(\alpha) = 0$ or $P_2(\neg\alpha) = 0$.

Suppose that $P_2(\circ\alpha) = 1$. If $P_2(\alpha) = 1$ and $P_2(\neg\alpha) = 1$, then Theorem 7, Item (2) leads to being absurd.

(Bival6) $P_2(\neg\neg\alpha) = 1 \iff P_2(\alpha) = 1$.

This item follows directly from axioms **Ax12** and **Ax13** and tautologicity.

(Bival7) $P_2(\neg \circ \alpha) = 1 \implies P_2(\alpha) = 1$ and $P_2(\neg\alpha) = 1$.

Suppose that $P_2(\neg \circ \alpha) = 1$. By Theorem 7, Item (5), $P_2(\circ\alpha) = 0$. Hence, by the same Theorem 7, Item (2), $P_2(\circ\alpha) = 0 = 2 - (P_2(\alpha) + P_2(\neg\alpha))$ and thus $P_2(\alpha) = 1$ and $P_2(\neg\alpha) = 1$.

□

Probabilistic semantics may be seen as an alternative to truth-valued semantics, with the intention to explain semantic notions, such as truth and consequence, in terms of probability functions. It has been shown in [16] that a semantics to standard logic (regarding soundness and completeness) can be provided without any appeal to model-theoretic or proof-theoretic concepts. The idea was later extended in [17], proving that a probabilistic semantics can be given to any extension of classical propositional logic. However, Theorem 8 is somewhat surprising, because the logic **Ci** is not an extension of classical propositional logic, just the contrary: although its language extends the language of classical logic, deductively, it is a contraction. Our (paraconsistent) probabilistic semantics to **Ci** shows, therefore, that an alternative to truth-valued semantics in terms of non-standard probability functions can be provided even in cases of the contractions of classical logic.

4. Conditional Probabilities and Paraconsistent Updating

Perhaps the most interesting use of probability in paraconsistent logic is to help the so-called Bayesian epistemology or the formal representation of belief degrees in philosophy. The well-known Bayes' rule permits one to update probabilities as new information is acquired and, in the paraconsistent case, even when such new information involves some degree of contradictoriness.

We define, as usual, the conditional probability of α given β for $P(\beta) \neq 0$ as:

$$P(\alpha/\beta) = \frac{P(\alpha \wedge \beta)}{P(\beta)}$$

The traditional Bayes' theorem for conditionalization says, for $P(\beta) \neq 0$:

$$P(\alpha/\beta) = \frac{P(\beta/\alpha) \cdot P(\alpha)}{P(\beta)}$$

As usual, $P(\alpha)$ here denotes the prior probability, i.e, is the probability of α before β has been observed. $P(\alpha/\beta)$ denotes the posterior probability, i.e., the probability of α after β is observed. $P(\beta/\alpha)$ is the likelihood or the probability of observing β given α, and $P(\beta)$ is called the marginal likelihood or "model evidence".

We can now set up a paraconsistent version of Bayes' theorem, as suggested in [18], by making the marginal likelihood $P(\beta)$ to be analyzed in terms of $P(\alpha)$, $P(\neg \alpha)$ and $P(\alpha \wedge \neg \alpha)$ when $P(\alpha \wedge \neg \alpha) \neq 0$.

It is convenient to show at this point a simple, yet pivotal generalization of the classical theorem of total probability:

Theorem 9. *Theorem of total paraconsistent probability:*
$P(\beta) = P(\beta \wedge \alpha) + P(\beta \wedge \neg \alpha) - P(\beta \wedge \alpha \wedge \neg \alpha)$

Proof. Since $\vdash_{\mathbf{Ci}} \alpha \vee \neg \alpha$, then $\beta \vdash_{\mathbf{Ci}} (\beta \wedge (\alpha \vee \neg \alpha)$, and clearly (by **Ax4**) $\beta \wedge (\alpha \vee \neg \alpha) \vdash_{\mathbf{Ci}} \beta$; hence, by Theorem 6:

$$P(\beta) = P(\beta \wedge (\alpha \vee \neg \alpha)).$$

Therefore,

$$P(\beta) = P(\beta \wedge (\alpha \vee \neg \alpha)) = P((\beta \wedge \alpha) \vee (\beta \wedge \neg \alpha)).$$

On the other hand, by the axiom of tautologicity (Definition 3):

$$P((\beta \wedge \alpha) \vee (\beta \wedge \neg \alpha)) + P((\beta \wedge \alpha) \wedge (\beta \wedge \neg \alpha)) = P(\beta \wedge \alpha) + P(\beta \wedge \neg \alpha)$$

hence:

$$P(\beta) + P((\beta \wedge \alpha) \wedge (\beta \wedge \neg \alpha)) = P(\beta \wedge \alpha) + P(\beta \wedge \neg \alpha)$$

and since:

$$P((\beta \wedge \alpha) \wedge (\beta \wedge \neg \alpha)) = P(\beta \wedge \alpha \wedge \neg \alpha)$$

it follows that:

$$P(\beta) = P(\beta \wedge \alpha) + P(\beta \wedge \neg \alpha) - P(\beta \wedge \alpha \wedge \neg \alpha).$$

□

Theorem 10. *Paraconsistent Bayes' Conditionalization Rule (PBCR): If* $P(\alpha \wedge \neg \alpha) \neq 0$*, then:*

$$P(\alpha/\beta) = \frac{P(\beta/\alpha) \cdot P(\alpha)}{P(\beta/\alpha) \cdot P(\alpha) + P(\beta/\neg \alpha) \cdot P(\neg \alpha) - P(\beta/\alpha \wedge \neg \alpha) \cdot P(\alpha \wedge \neg \alpha)}$$

Proof. First notice that $P(\alpha \wedge \neg\alpha) \neq 0$ entails that $P(\alpha) \neq 0$ and $P(\neg\alpha) \neq 0$ in view of Definition 3 and **Ax4**, so the quotient is well defined. Suppose we have two contradictory hypothesis, α and $\neg\alpha$, and wish to compute the probability of α based on evidence β. Since the definition of conditional probability gives:

$$P(\alpha/\beta) = \frac{P(\beta/\alpha) \cdot P(\alpha)}{P(\beta)}$$

it remains to compute the marginal likelihood $P(\beta)$ depending on $P(\alpha)$ and $P(\neg\alpha)$. This follows immediately by Theorem 9, dividing and multiplying each term by, respectively, $P(\alpha)$, $P(\neg\alpha)$ and $P(\alpha \wedge \neg\alpha)$ (which are not zero). □

It is clear that this rule generalizes the classical conditionalization rule, as it reduces to the classical case if $P(\alpha \wedge \neg\alpha) = 0$ or if α is consistent: indeed, in the last case, $P(\beta \wedge \circ\alpha) = P(\beta \wedge \circ\alpha \wedge \alpha) + P(\beta \wedge \circ\alpha \wedge \neg\alpha)$ since $P(\circ\alpha \wedge \alpha \wedge \neg\alpha) = 0$.

As a slogan, we could summarize PBCR as saying: "Posterior probability is proportional to likelihood times prior probability and inversely proportional to the marginal likelihood analyzed in terms of its components". It is possible, however, to formulate other kinds of conditionalization rules by combining the notions of conditional probability, contradictoriness, consistency and inconsistency.

Example 1. *As an example, suppose that a doping test for an illegal drug is such that it is 98% accurate in the case of a regular user of that drug (i.e., it produces a positive result, showing "doping", with probability 0.98 in the case that the tested individual often uses the drug), and 90% accurate in the case of a non-user of the drug (i.e., it produces a negative result, showing "no doping", with probability 0.9 in the case that the tested individual has never used the drug or does not often use the drug).*

Suppose, additionally, that: (i) it is known that 10% of the entire population of all athletes often uses this drug; (ii) that 95% of the entire population of all athletes does not often use the drug or has never used it; and (iii) that the test produces a positive result, showing "doping", with probability 0.11 for the whole population, independent of the tested individual.

Let us use the following abbreviations mnemonically:

D: the event that the drug test has declared "doping" (positive) for an individual;

C: the event that the drug test has declared "clear" or "no doping" (negative) for an individual;

A: the event that the person tested often uses the drug;

¬A: the event that the person tested does not often use the drug or has never used it.

We know that $P(A) = 0.1$ and $P(\neg A) = 0.95$. However, the situation is contradictory with respect to the events A and $\neg A$, as they are not excludent. Therefore, by finite additivity, $P(A \vee \neg A) = 1 = (P(A) + P(\neg A)) - P(A \wedge \neg A)$, and thus, $P(A \wedge \neg A) = (P(A) + P(\neg A)) - 1 = 0.05$

Furthermore, as given in the problem, $P(D/A) = 0.98$, $P(C/\neg A) = 0.9$ and $P(D) = 0.11$. The results of the test have no paraconsistent character, since the events D ('doping') and C ('no doping') exclude each other. Thus, $P(D/\neg A) = 1 - P(C/\neg A) = 0.1$ and $P(C/A) = 1 - P(D/A) = 0.02$.

Suppose someone has been tested, and the test is positive ("doping"). What is the probability that the tested individual regularly uses this illegal drug, that is what is $P(A/D)$?

By applying the paraconsistent Bayes' rule:

$$P(A/D) = \frac{P(D/A) \cdot P(A)}{P(D/A) \cdot P(A) + P(D/\neg A) \cdot P(\neg A) - P(D/A \wedge \neg A) \cdot P(A \wedge \neg A)}$$

since $P(A \wedge \neg A) \neq 0$.

All of the values are known, with the exception of $P(D/A \wedge \neg A)$. Since:

$$P(D/A \wedge \neg A) = \frac{P(D \wedge A \wedge \neg A)}{P(A \wedge \neg A)}$$

it remains to compute $P(D \wedge A \wedge \neg A)$. It follows directly from Theorem 9 that $P(D \wedge A \wedge \neg A) = P(D \wedge A) + P(D \wedge \neg A) - P(D) = P(D/A).P(A) + P(D/\neg A).P(\neg A) - P(D) = 0.083$. Therefore, by plugging in all of the values, it follows that $P(A/D) = 35.5\%$.

The probability $P(\neg A/D)$ that a tested individual never uses the drug or just uses it sporadically, given that the test has been positive, analogously computed via the paraconsistent Bayes' rule, is 34.4%.

As shown below, this example suggests that the paraconsistent Bayes' conditionalization rule is more sensitive to the test parameters than traditional conditionalization. The following tables compare the paraconsistent results with the results obtained by trying to remove the contradiction involving the events A (the event that the person tested often uses the drug) and $\neg A$ (the event that the person tested does not often use the drug or has never used it), that is by trying to make them "classical". The two tables refer to two kinds of tests: a less reliable test, with 10% of false positives (Table 1), and a more reliable test, with 2% of false positives (Table 2).

Since A and $\neg A$ overlap by 5%, we might think, thus, about reviewing the values, by removing the overlap according to three hypothetical scenarios: an alarming scenario, by lowering the value of $\neg A$ by 5%; a happy scenario, by lowering the value of A by 5%; and a cautious scenario, by dividing the surplus equally between A and $\neg A$ and computing the probability $P(A/D)$ that the tested individual regularly uses this illegal drug.

Table 1. Less reliable test.

Alarming Scenario	Cautious Scenario	Happy Scenario
$P(A) = 10\%$	$P(A) = 7.5\%$	$P(A) = 5\%$
$P(\neg A) = 90\%$	$P(\neg A) = 92.5\%$	$P(\neg A) = 95\%$
$P(D/A) = 98\%$	$P(D/A) = 98\%$	$P(D/A) = 98\%$
$P(D/\neg A) = 10\%$	$P(D/\neg A) = 10\%$	$P(D/\neg A) = 10\%$
Result	**Result**	**Result**
$P(A/D) = 52\%$	$P(A/D) = 44\%$	$P(A/D) = 34\%$

Using paraconsistent probabilities, however, one obtains a lower value concerning this less reliable test, namely, $P(A/D) = 35.5\%$, tending towards the "happy" hypothetical scenario where fewer people use the drug.

Table 2. Less reliable test.

Alarming Scenario	Cautious Scenario	Happy Scenario
$P(A) = 10\%$	$P(A) = 7.5\%$	$P(A) = 5\%$
$P(\neg A) = 90\%$	$P(\neg A) = 92.5\%$	$P(\neg A) = 95\%$
$P(D/A) = 98\%$	$P(D/A) = 98\%$	$P(D/A) = 98\%$
$P(D/\neg A) = 2\%$	$P(D/\neg A) = 2\%$	$P(D/\neg A) = 2\%$
Result	**Result**	**Result**
$P(A/D) = 84\%$	$P(A/D) = 80\%$	$P(A/D) = 72\%$

In this second, more reliable test, by computing the result directly via paraconsistent probabilities, one obtains a higher value, namely $P(A/D) = 79\%$, tending towards the "cautious" hypothetical scenario.

The values $P(D/A)$ and $P(C/\neg A)$ are known, respectively, as sensitivity and specificity, and the positive likelihood ratio is defined as *sensitivity*$/1 - $*specificity* (similarly, the negative likelihood ratio is defined as $1 - $*sensitivity*$/$*specificity*. Defining such measures for paraconsistent probabilities would help to assess their meaning, but this kind of approach is not the intention of this paper.

These simple examples suggest the following interpretation about Bayesian paraconsistent updates: When the test is less reliable, paraconsistent probabilities tend toward cautious optimism:

values tend to reflect the most favorable outcome. On the other hand, as the test gets more reliable, paraconsistent probabilities tend toward cautious realism, in the sense of favoring more realistic expectations of undesirable outcomes. This notwithstanding, the test is cautious in all cases.

Such numerical results are very much connected to the philosophical profile of paraconsistent logics, which naturally endorses cautious reasoning about contradictions: when you find a contradiction, it is better to carefully analyze its causes, instead of risking throwing the baby out with the bath water.

It is worth noting that, apparently, our comparisons above would seem to be futile. Since, as an unwary reader could think, standard probabilistic merging techniques would be directly applied to define a belief function consensus (as studied, e.g., in [19] and further extended in several ways), this is however highly debatable, as discussed in [20] by one of the founders of Dempster–Shafer theory, who argues against global consensus methods and in favor of direct reassessment of the items of evidence that contribute to a belief function.

Besides such a debate of whether or not a consensus of belief functions can be directly computed, it is not obvious at all how to understand beliefs in terms of probability theory. As put by J. Y. Halpern and R. Fagin in [21], page 3:

> If we view belief as generalized probability, then it makes sense to *update* beliefs but not *combine* them. If we view beliefs as a representation of evidence, then it makes sense to combine them, but not update them. This suggests that the rule of combination is appropriate only when we view beliefs as representations of evidence

The belief functions defined by our probability functions, of course, are generalized probabilities, and we align with Halpern and Fagin, among others, in that it makes sense to update beliefs instead of combining them.

Although this first example is just a suggestive sample of what can be done with a robust calculus of probabilities based on a well-founded paraconsistent logic, it is convenient to recall that false positives themselves can be regarded as contradictory when false positive results are more probable than true positive tests; this occurs when the incidence of a certain condition (for instance, disease) in a population is below the probability of a false positive rate. In this case, the traditional Bayes' conditionalization rule already incorporates a mechanism for rationally handling some types of contradictory data without falling into trivialization. Indeed, in [22], the paraconsistent paradigm is invoked as a tool to evaluate the sensitivity of the traditional Full Bayesian Significance Test (FBST) value regarding changes in the prior or reference density. That paper argues that such an intuitive measure of inconsistency can be made rigorous in the context of paraconsistent logic and bilattice structures. Our intention is different, as we start from a logical-paraconsistent approach, but both views can be, in principle, combined.

Furthermore, in [23], a conviction measure and a loss function are defined, intended to be used for evaluating financial operations strategies. A classification tree learning algorithm is thus defined over such a conviction measure, with the advantage of outputting more cautious decisions. It is not implausible that our generalization of Bayes' updating procedure could be used with similar applications in mind, but this is a subject of further investigation.

5. Paraconsistent Probability Spaces

As is historically recognized (see, e.g., [24]), two main competing schools of probability emerged in Europe in the 17th century, leading to different methods of statistical inference and estimation, frequentist (based on the laws of large numbers and concerned with stochastic laws of chance) and epistemological (refereeing to credence or reasonable degrees of belief, somehow connected to the modern Bayesianism). However, there is also a second competition, represented by the theories of logical probability (in contemporary times, Keynes/Carnap) versus measure-theoretical approaches to

probability (Kolmogorov). These pair of bipartite traditions are not unrelated, but nobody knows for sure how all this is related to the debate between logic and probability (see below).

The reference to probability on sentences is old and comes from Gottfried W. Leibnitz (1646–1716), Jacobus Bernoulli (1654–1705), Johann Heinrich Lambert (1728–1777), Bernard Bolzano (1781–1848), Augustus De Morgan (1806–1871) and George Boole (1815–1864) and, in the 19th and 20th century John Venn (1834–1923), Hugh MacColl (1837–1909), Charles S. Pierce (1839–1887), Hans Reichenbach (1891–1953), John M. Keynes (1883–1946) and Rudolph Carnap (1891–1970), just to stay with some famous names.

On the other hand, the reference to probability on sets is more recent. Kolmogorov introduced in his classic book in German of 1933, translated as [25], what he called the 'elementary theory of probability', which nowadays is widely used by mathematicians, statisticians and engineers and connected to measure theory. Probability on sentences is more common among philosophers and logicians.

Probability can, alternatively, be based on game theory rather than measure theory, as developed in [26], viewing probability as a perfect information game between two players. The so-called axiom of continuity, considered not to be well motivated, is not assumed in [26]. As put in [27]:

> Countable additivity for probability has always been controversial. Émile Borel, who introduced it, and Andrei Kolmogorov, who confirmed its role in measure-theoretic probability, were both ambivalent about it. They saw no conceptual argument for requiring probabilities to be countably additive. It is merely mathematically convenient to assume they are. As Kolmogorov explained in his *Grundbegriffe*, countable additivity has no meaning for empirical experience, which is always finite, but it is mathematically useful. We can elaborate Kolmogorov's explanation by pointing out that infinities enter into applied mathematics not as representation but as simplification.

Probability on sets (more properly called probability functions) assigns probabilities directly to sets. Although events, statements or predicates in many cases can be expressed as sets, these two accounts are not the same, although they are inherently equivalent in the classical scenario. In a paper that surveys and discusses K. Popper's contributions to the theory of probability, H. Leblanc in [28] recalls that Popper reacted against the dependency of probability theory on the Boolean algebra of sets and proposed his notion of absolute probability motivated by this criticism.

The equivalence between probability on sentences and probability on sets is far from obvious under the paraconsistent paradigm, and our proposal would not be complete if we could not offer a similar approach from the paraconsistent perspective.

Definition 4. *A paraconsistent probability space is a structure* $\langle \Omega, \Sigma, \Pi, P_\mu \rangle$ *where:*

1. Ω *is the sample space composed of all possible outcomes*
2. $\Sigma \subseteq \wp(\Omega)$ *is a set of events, such that* Σ *is a* $\sigma^\star$-*algebra, i.e.:*

    (a) $\varnothing \in \Sigma$, $\Omega \in \Sigma$ *and* $\Pi \in \Sigma$;
    (b) Σ *is closed under* $\cup, \cap$ *and countable unions;*
    (c) Σ *is closed under the following two binary operations:*

        i. $X^\star : \Sigma \mapsto \Sigma$, *such that* $\bar{X} \subseteq X^\star$, *where* $\bar{X}$ *is the usual complement.*
        ii. $\bigcirc X : \Sigma \mapsto \Sigma$, *such that* $\bigcirc X \cap X^\star = \bar{X}$
    (d) Π *is the set of all consistent outcomes;*

3. *The map* $P_\mu : \Sigma \mapsto [0, 1]$ *is a probability measure satisfying the following conditions:*

    (a) $P_\mu(\Omega) = 1$ *and* $P_\mu(\varnothing) = 0$
    (b) *If* $S_1, \cdots S_n \in \Sigma$ *are pairwise disjoint, then* $P_\mu(\bigcup_{i=1}^{\infty} S_i) = \sum_{i=1}^{\infty} P_\mu(S_i)$

The point of assigning a probability interpretation to logic systems is how legitimate it is to regard probability as attached to logical statements, instead of directly to events. This duality causes

perplexity to some, since statisticians and engineers may find confusing or unnecessary this passage to logic and would prefer to deal with events directly (in this case, the notion of a random variable is essential). Philosophers and logicians, on the other hand, may find it natural to attach probabilities to statements, not to events (and in this case, the notion of consequence relation is essential).

A (classical) probability space is just a particular case of paraconsistent probability space $\langle \Omega, \Sigma, P \rangle$ where Π is empty and the operation $\bigcirc X : \Sigma \mapsto \Sigma$ is the identity on Σ (and consequently $X^\star = \bar{X}$).

In the classical case, a definite link between probability on events and probability on sentences (in the cases of finite additivity) can be defined so that the probability measures of information in both cases is the same. For the cases that demand infinite additivity, an infinitary propositional language or a first-order logic is required (see [29], pages 401–404). For De Finneti (see [30]), as much as for Borel and Kolmogorov, infinite additivity is just a question of mathematical convenience, not strictly justified by the concept of probability:

> Its success owes much to the mathematical convenience of making the calculus of probability merely a translation of modern measure theory. [...] No-one has given a real justification of countable additivity (other than just taking it as a 'natural extension' of finite additivity).

In the cases where Σ is closed under $\cup$, $\cap$ and finite unions, a $\sigma^\star$-algebra is referred to as a σ-algebra.

The connection between probability on events and probability on sentences in the classical case is granted through algebraic maneuvers à la Lindenbaum algebra, by showing that probability measures on sentences turn out to be individual cases of probability measures on events. The argument is essentially the following: denote the set of classical bivaluations in its language L as *Val*, and assign to each sentence $\varphi \in L$ a subset of *Val*, defined by $[[\varphi]] = \{v \in Val | v(\varphi) = 1\}$.

It can be inductively proven that:

1. $[[\varphi \vee \psi]] = [[\varphi]] \cup [[\psi]]$;
2. $[[\varphi \wedge \psi]] = [[\varphi]] \cap [[\psi]]$;
3. $[[\neg \varphi]] = Val - [[\varphi]]$.

Therefore, the universe $A = \{[[\varphi]] \in \wp(Val) | \varphi \in L\}$ acquires a σ-algebra structure. Now, given any probability distribution $P : L \mapsto [0,1]$ as in Definition 3, one defines a probability measure over sets $P_\mu : A \mapsto [0,1]$, as in Definition 4, by $P_\mu([[\varphi]]) = P(\varphi)$, for any $\varphi \in L$. Clearly P_μ is well-defined in view of Theorem 6, Item (2).

The converse can be easily determined in the cases of finite probability spaces $\langle \Omega, P_\mu \rangle$ representing the outcomes of a certain stochastic experiment $\Omega = \{\omega_1, \cdots, \omega_n\}$ (say, throwing dice). In such cases, one can define a corresponding probabilist logic by setting atomic variables $\alpha_1, \cdots, \alpha_n$, where α_i means that ω_i is the outcome of the experiment, and sentences $\rho_i = \neg\alpha_1 \wedge \cdots \wedge \neg\alpha_{i-1} \wedge \alpha_i \wedge \neg\alpha_{i+1} \wedge \cdots \wedge \neg\alpha_n$, saying that ω_i occurs independently, and defining $P_L(\rho_i) = P(\{\omega_i\})$ for $i = 1, \cdots, n$ and $P_L(\lambda) = 0$ for any other sentences in these atomic variables. It is easy to see that $\langle L, P_L \rangle$ is a propositional probability model. Some more details can be found, for the classical case, in [31], Chapter 4.7. A concrete example of a paraconsistent probability space is given by the structure:

Example 2. $\langle \Omega, \Sigma, \Pi, P_\mu \rangle$ *where:*

1. Ω *is any set (representing all possible outcomes)*
2. $\Sigma \subseteq \wp(\Omega)$ *is the set of all events, such that Σ is a σ-algebra, i.e.:*

   (a) $\varnothing \in \Sigma$, $\Omega \in \Sigma$ *and* $\Pi \in \Sigma$;
   (b) Σ *is closed under* $\cup$ *and* $\cap$;
   (c) Σ *is closed under the binary operations:*

       i. $X^\star = \bar{X} \cup \bar{\Pi}$

    ii. $\bigcirc X = \Pi$

In this case, Π represents the consistent events (from the point of view of the logic **Ci**). When $\Pi = \Sigma$, this kind of paraconsistent probability space turns out to be a classical probability space. The structure $\langle \Omega, \Sigma, \Pi \rangle$ is a particular case of a paraconsistent algebra of sets, as investigated in [32], and the results therein (particularly Theorems 4 and 5) can be adapted to give a precise connection between the concepts of paraconsistent probability on events and probability on sentences in the **Ci** logic (again, in the cases of finite additivity).

This kind of structure, of interest for mathematical statistics, would permit dealing with paraconsistent probability distributions in the way that is familiar for applications.

The main differences between 'paraconsistent probability logics' and 'paraconsistent probability spaces' are connected to the fact that the former are concerned with the transmission of probability through valid inferences, while the latter concentrate on measures of sets, the distribution of probabilities and their consequences, permitting one to treat random variables, expectation, central limit theorem, etc. Of course, as much as in the classical case, philosophical issues concern probabilistic logics, not spaces.

Probability spaces and probability logic, as we have seen, make for equivalent views in the classical case (at least for the finite situations), and a similar proof can be adapted for our paraconsistent probability logic and paraconsistent probability spaces (keeping in mind that in both cases, the definitions reflect the inherent nature of or underlying logic **Ci**). A deep investigation about paraconsistent probability spaces in the direction of a more sophisticated treatment involving random variables, measure theory and similar topics, however, is out of the scope of the present paper.

6. Discussion: From Paraconsistent Probability to Paraconsistent Possibility

Possibility theory is a theory of uncertainty used in areas, such as artificial intelligence, non-monotonic reasoning, belief revision and similar domains, to express uncertain knowledge in scenarios of incomplete information. It is regarded as a kind of imprecise probability theory, at times seen as a generalization of probability, sometimes as its rival (cf., e.g., [33]). Possibility theory uses a pair of dual set functions (possibility and necessity measures) instead of only one measure, in this way differing from probability.

As a formalism to model and reason with uncertain information, possibility theory is also connected to the Dempster–Shafer theory of evidence (see [34]). Probability, possibility and other credal calculi are alternative formalisms, with distinct motivations, and sometimes, it may be important to combine different belief measures for better solving certain problems, as explained in [35].

Possibility theory can be easily and naturally extended over LFIs, by defining possibility and necessity functions (as, e.g., in [36] and especially in [37], where the notion of degrees of support is investigated as generalizations of degrees of belief) simulating our paraconsistent probability measures over LFIs. A deeper study on the meaning and applications of paraconsistent possibility theory, as well as its inter-relations to paraconsistent probability are postponed to further work.

7. Summary, Comments and Conclusions

We have reviewed some basic points about paraconsistency, characterizing a logic of formal inconsistency as a paraconsistent logic endowed with a notion of consistency $\circ$ and a negation $\neg$, which is free from trivialism. This means that a contradiction expressed by means of the negation $\neg$ does not necessarily trivialize the underlying consequence relation, although consistent contradictions do explode. We then defined a measure of probability to one of such logics, the system **Ci**, taking profit from the underlying notion of consistency and essaying the first steps towards paraconsistent Bayesian updating.

One of the most important topic in statistics is how to ensure reliable uncertain inferences. If we agree with this, the notion of probability and its connection to logic are of fundamental

importance. Of course, it is also possible to propose a modal paraconsistent approach to probability based on paraconsistent modalities (see [38]), following the lines of [39], but this has to wait for a better motivation.

How can we interpret paraconsistent probabilities? One possible viewpoint is to interpret paraconsistent probabilities as degrees of belief that a rational agent attaches to events, in such a way that such degrees respect the following principles: the necessary events (for instance, tautologies) get the maximum degrees; impossible events (for instance, bottom particles) get the lowest degrees; probabilities respect logical consequence; and finite additivity is guaranteed (that is, conjunction and disjunction retain their classical interpretation). The last condition seems to be less obvious, but the Dutch book argument provides, at least in the classical case, a line of justification for keeping finite additivity. A Dutch book is a sequence of bets, regarded by the agent as fair, but that in the long range guarantee the agent's loss. De Finetti proved in [40] that a Dutch book can be constructed against an agent whose degrees of belief do not respect finite additivity probabilities.

J. B. Paris in [41] has shown that a generalization of the standard Dutch book theorem (his Theorem 5) applies to several non-classical logics, among them to paraconsistent logics where the meaning of conjunction and disjunction satisfies the clause for finite additivity in Definition 3. This generalization encompasses our notion of paraconsistent probability based on the logic **Ci** as well as paraconsistent probability theories based on several other LFIs. More recently, it has been shown in [42] that the Dutch book argument can be further extended to the domain of MV-algebras, providing a logical characterization of coherence for imprecise probability.

It should be taken into account that our underlying logic **Ci** enlarges the classical scenarios in significant ways: so for instance, even if impossible events should have degree zero by a rational agent, neither such events are necessarily unlikely nor a contradiction is an impossible event (although a consistent contradiction, as commented above, is impossible).

The question of how intimate is the relationship between logic and probability has been a controversial issue for more than three centuries. For instance, F. Ramsey considered in [9] the theory of probability as a branch of logic, in which we find echoes of Leibniz and his defense of a "new kind of logic" devoted to questions of jurisprudence. Leibniz believed that mathematical calculations could help the legal standpoint, but he lacked the relevant combinatorial tools. Boole also presented a calculus of probabilities, based on his logical calculus, developed in the earlier part of his book *The Laws of Thought*. All this led to a tradition of regarding probabilities as a generalization of two-valued logic, as R. Jeffrey argues in [43].

Perhaps the two most relevant questions concerning the relationship between logic and probability are: (i) whether the laws of probability should be classified as laws of logic; and (ii) how logic and probability could be combined to refine reasoning. The interpretation of probability has been a torrid controversial question. If we see consequence relations as preserving degrees of probability instead of truth (as in standard logic), then we gain an interpretation that sees probability as part of logic. We do not intend to enter into this thorny issue in this paper. What concerns us here is that, independent of the chosen interpretation, logic and probability can be combined in a mathematically and philosophically well-reasoned way.

In this respect, considering that probability theory differs from classical logic in various aspects, that paraconsistent logic differs from the classical stance, as well, and that both are tolerant to contradictions, inexactness, and so on, their combination offers a new and exciting reasoning paradigm.

Some specific problems connected to probabilities in non-classical logics are pointed out by J. R. G. Williams in [14]:

> Studying axiomatizations of non-classical probabilities is an open-ended task. Can we extend the results of Paris, Mundici *et al.*, and get more general sense of what set of axioms are sufficient to characterize expectations of truth value? A major obstacle here is the appeal throughout to the additivity principle (P3) and its variants, which is the only one to turn

essentially on the behavior of particular connectives. Is there a way of capturing its content in terms of logical relations between sentences rather such hardwired constraints?

Our axiomatics does not suffer from such problems and naturally generalizes the additivity principle. Williams also warns (on p. 17) that the generalization of conditional probability in some cases does not guarantee that $P(\alpha|\alpha) = 1$. Again, this is not a problem in our setting, and yet, another possible source of problems, the fact that $P(\alpha \vee \beta) = P(\alpha) + P(\beta)$ should hold for all logically-incompatible α and β, for instance, is treated in our Theorem 7.

As concerns axiomatizations of non-classical probabilities, some other LFIs are natural candidates to be used in connection with probabilities. One of the most promising is the three-valued paraconsistent logic **LFI1** (see [44]): it is maximal with respect to classical logic (thus, in a certain sense closer to classical logic), enjoys some forms of De Morgan laws, and, above all, is algebraizable. This task is beyond the scope of this paper and will be postponed to future work.

De Finetti provocatively, in [30]), said that probability does not exist, meaning that probability exists only subjectively, within the individual minds, to which Leibniz would certainly agree. For de Finetti, probability does not need to exist as a property of the outside world and, thus, is a purely epistemic concept. In this sense, paraconsistent probability as attached to the logics of formal inconsistency can be seen as a measure of the persistence of belief under contradictions within the individual minds and has epistemological roots with strong connections with the notion of evidence, as defended in [2].

Many arguments maintain that credences should conform to the probability calculus, but since of course an agent can hold contradictory credences without losing its status of being ideally rational (and hopefully, immune to a Dutch book argument), its rationality should be coherent to a paraconsistent probability calculus. It seems to us that what we have in hand is a promising tool to enlarge the notion of probability.

Author Contributions: Both authors have equally contributed to the theoretical aspects of this work, as well as to the conception and design of the numerical experiments and their interpretation. Both authors have read and approved the final manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Leibniz, G.W. *Sämtliche Schriften und Briefe. Reihe 6: Philosophische Schriften. Band 2: 1663–1672*; Akademie Verlag: Berlin, Germany, 1966; p. 492. (In German)
2. Carnielli, W.; Rodrigues, A. What Contradictions Say (and What They Say Not). *CLE e-Prints* **2012**, *12*. Available online: ftp://logica.cle.unicamp.br/pub/e-prints/vol.12,n.2,2012.pdf (accessed on 23 August 2016).
3. Carnielli, W. The single-minded pursuit of consistency and its weakness. *Stud. Log.* **2011**, *97*, 81–100.
4. Carnielli, W.; Coniglio, M.E.; Marcos, J. Logics of formal inconsistency. In *Handbook of Philosophical Logic*, 2nd ed.; Gabbay, D.M., Guenthner, F., Eds.; Springer: Amsterdam, The Netherlands, 2007; Volume 14, pp. 1–93.
5. Carnielli, W.; Coniglio, M.E. *Paraconsistent Logic: Consistency, Contradiction and Negation*; Logic, Epistemology, and the Unity of Science Series 40; Springer: Cham, Switzerland, 2016.
6. Williams, J.N. Inconsistency and contradiction. *Mind* **1981**, *90*, 600–602.
7. Carnielli, W.; Coniglio, M.E.; Podiacki, R.; Rodrígues, T. On the way to a wider model theory: Completeness theorems for first-order logics of formal inconsistency. *Rev. Symb. Log.* **2015**, *3*, 548–578.

8. Bueno-Soler, J.; Carnielli, W. Experimenting with Consistency. Available online: http://www.cle.unicamp. br/e-prints/vol_14,n_3,2014.html (accessed on 23 August 2016).

9. Ramsey, F.P. Truth and Probability (1926). In *Studies in Subjective Probability*; Kyburg, H.E., Smokler, H.E., Eds.; Wiley: New York, NY, USA, 1964; pp. 61–92.

10. Mares, E. Paraconsistent probability theory and paraconsistent Bayesianism. *Log. Anal.* **1997**, *160*, 375–384.

11. Priest, G. *In Contradiction: A Study of the Transconsistent*; Martinus Nijhoff: Amsterdam, The Netherlands, 1987.

12. Roberts, L. Maybe, maybe not: Probabilistic semantics for two paraconsistent logic. In *Frontiers of Paraconsistent Logic*, Proceedings of the First World Congress on Paraconsistency, Ghent, Belgium, 30 July–2 August 1997; Batens, D., Mortensen, C., Priest, G., van Bendegem, J.P., Eds.; Logic and Computation Series; Research Studies Press: Baldock, UK, 2000; pp. 233–254.

13. Mundici, D. Bookmaking over infinite-valued events. *Int. J. Approx. Reason.* **2006**, *43*, 223–240.

14. Williams, J.R.G. Probability and Non-Classical Logic. Available online: http://www.personal.leeds.ac.uk/ ~phljrgw/wip/nonclassicalprobabilitypostrevisions.pdf (accessed on 23 August 2016).

15. Weatherson, B. From Classical to Intuitionistic Probability. *Notre Dame J. Form. Log.* **2003**, *44*, 111–123.

16. Field, H. Logic, meaning, and conceptual role. *J. Philos.* **1977**, *84*, 379–409.

17. Morgan, C.G. There is a probabilistic semantics for every extension of classical sentence logic. *J. Philos. Log.* **1981**, *11*, 431–442.

18. Carnielli, W. Uma lógica da modalidade econômica? *Rev. Bras. Filos.* **2009**, *232*, 1–18. (In Portuguese)

19. Wagner, C.G. Consensus for belief functions and related uncertainty measures. *Theory Decis.* **1989**, *26*, 295–304.

20. Shafer, G. [Combining Probability Distributions: A Critique and an Annotated Bibliography]: Comment. *Stat. Sci.* **1986**, *1*, 135–137.

21. Halpern, J.Y.; Fagin, R. Two views of belief: Belief as generalized probability and belief as evidence. *Artif. Intell.* **1992**, *54*, 275–317.

22. Stern, J.M. Paraconsistent Sensitivity Analysis for Bayesian Significance Tests. *Advances in Artificial Intelligence —SBIA 2004*; Lecture Notes in Computer Science Series 3171; Bazzan, A.L.C., Labidi, S., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; pp. 134–143.

23. Lauretto, M.S.; Nakano, F.; Ribeiro, C.O.; Stern, J.M. Real attribute learning algorithm. In Proceedings of the International Institute of Informatics and Systems, Anaiysis and Synthesis, Orlando, FL, USA, 12–16 July 1998; pp. 315–321.

24. Hacking, I. *The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference*; Cambridge University Press: Cambridge, UK, 2006.

25. Kolmogorov, A.N. *Foundations of the Theory of Probability*; Chelsea: New York, NY, USA, 1956. Translation of Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer: Berlin, Germany, 1933.

26. Shafer, G.; Vovk, V. *Probability and Finance, It's Only a Game!* Wiley-Interscience: New York, NY, USA, 2001.

27. Shafer, G.; Vovk, V. Question 1. Continuity. Available online: http://www.probabilityandfinance.com/ faq/question1.pdf (accessed on 23 August 2016).

28. Leblanc, H. Popper's formal contributions to probability theory. In *Perspectives on Psychologism*; Notturno, M.A., Ed.; Brill: Leiden, The Netherlands, 1989; pp. 341–367.

29. Williamson, J. Probability logic. In *Handbook of the Logic of Argument and Inference: The Turn Towards the Practical*; Gabbay, D.M., Johnson, R.H., Ohlbach, H.J., Woods, J., Eds.; Elsevier: Amsterdam, The Netherlands, 2002; pp. 397–424.

30. De Finetti, B. *Theory of Probability: A Critical Introductory Treatment*; Wiley: Queensland, Australia, 1974.

31. Hailperin, T. *Sentential Probability Logic: Origins, Development, Current Status, and Technical Applications*; Lehigh University Press: Bethlehm, PA, USA, 1996.

32. Carnielli, W.A.; de Alcantara, L.P. Paraconsistent algebras. *Stud. Log.* **1984**, *43*, 79–88.

33. Dubois, D.; Prade, H. Possibility Theory, Probability Theory and Multiple-valued Logics: A Clarification. *Ann. Mathe. Artif. Intell.* **2001**, *32*, 35–66.

34. Dubois, D.; Prade, H. Representation and combination of uncertainty with belief functions and possibility measures. *Comput. Intell.* **1988**, *4*, 244–264.

35. Stern, J.M.; Pereira, C.A.B. Bayesian epistemic values: Focus on surprise, measure probability! *Log. J. IGPL* **2014**, *22*, 236–254.

36. Besnard, P.; Lang, J. Possibility and necessity functions over non-classical logics. In Proceedings of the Tenth International Conference on Uncertainty in Artificial Intelligence, Seattle, WA, USA, 29–31 July 1994; Morgan Kaufmann Publishers: San Francisco, CA, USA, 1994; pp. 69–76.

37. Darwiche, A.Y.; Ginsberg, M.L. A symbolic generalization of probability theory. In Proceedings of the 10th Conference of the American Association for Artificial Intelligence, San Jose, CA, USA, 12–16 July 1992; pp. 622–627.

38. Bueno-Soler, J. Two semantic approaches to paraconsistent modalities. *Log. Univ.* **2010**, *4*, 137–160.

39. Heifetz, A.; Mongin, P. The modal logic of probability. In Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge, Evanston, IL, USA, 22–24 July 1998; Gilboa, I., Ed.; Morgan Kaufmann: San Francisco, CA, USA, 1998; pp. 175–185.

40. De Finetti, B. La prévision: Ses lois logiques, ses sources subjectives. *Ann. Inst. Henri Poincar*é **1937**, *17*, 1–68. (In French)

41. Paris, J.B. A note on the Dutch Book method. In Proceedings of the Second International Symposium on Imprecise Probabilities and Their Applications, Ithaca, NY, USA, 26–29 June 2001; Cooman, G.D., Fine, T., Seidenfeld, T., Eds.; Shaker Publishing: Maastricht, The Netherlands; pp. 301–306.

42. Fedel, M.; Hosni, H.; Montagna, F. A logical characterization of coherence for imprecise probabilities. *Int. J. Approx. Reason.* **2011**, *52*, 1147–1170.

43. Jeffrey, R. *Probability and the Art of Judgement*; Cambridge University Press: Cambridge, UK, 1992.

44. Carnielli, W.; Marcos, J.; de Amo, S. Formal Inconsistency and Evolutionary Databases. *Log. Log. Philos.* **2000**, *8*, 115–152.