Available online at www.sciencedirect.com
SCIENCE DIRECT®

Journal of Applied Logic 3 (2005) 97–118

JOURNAL OF APPLIED LOGIC

www.elsevier.com/locate/jal

Paraconsistent Informational Logic

Paola Forcheri [a], Paolo Gentilini [a,b]

[a] *Istituto di Matematica Applicata e Tecnologie Informatiche, Consiglio Nazionale delle Ricerche Via De Marini 6, 16149 Genova, Italy*
[b] *Istituto Regionale di Ricerca Educativa (IRRE), Liguria, Via Lomellini 15, 16124 Genova, Italy*

Available online 21 August 2004

Abstract

We introduce a Paraconsistent Informational Logic that formalizes the idea of conjectures which are acceptable as to the quality and the variety of the information that they convey with respect to a given theory **T**, even if they are classically inconsistent with **T**. The work constitutes an extension of a previously developed Informational Logic for classical frameworks, where a new notion of logical entropy measure H on formulas and on proofs plays a central role.

Keywords: Informational logic; Paraconsistent logic; Logical entropy measures; Syntactic probability; Formal conjectures

1. Introduction

Informational Logic aims to build formal contexts suitable to approximate, from a proof-theoretic standpoint, the conjectures carried out by an epistemic agent **T**. We figure out that, to carry out this formulation of conjectures, the agent **T**, given a statement L, analyses the decisions about L produced by a powerful inferential environment $\boldsymbol{\Omega}$, external to **T**. Then, **T** gives an estimate of the possibility of stating L through its own inference, by comparing its inferential capabilities with that of $\boldsymbol{\Omega}$. On this basis, a **T**-acceptability grade to L is assigned, that expresses the closeness of L with respect to **T**. Such value is the result of the comparison between the theoretical information on L produced by the

E-mail addresses: forcheri@ge.imati.cnr.it (P. Forcheri), gentilini@irre.liguria.it, gentilini@ge.imati.cnr.it (P. Gentilini).

1570-8683/$ – see front matter
doi:10.1016/j.jal.2004.07.013

environment $\mathbf{\Omega}$ and the estimate of the possible theoretical information on L that can be constructed by $\mathbf{T}$. To develop a formal setting apt to represent this kind of reasoning, we work as follows. Let us suppose that a sentence L is proved by the deductive apparatus $ded(\mathbf{\Omega})$ by a proof P. Then, a formal conjecture based on L from the standpoint of $\mathbf{T}$ is given by a pair $(L, p(L))$ where $p(L)$ is a measure of the probability of L to be provable using $ded(\mathbf{T})$. $p(L)$ is the result of a syntactic comparison between the real proof P of L in $\mathbf{\Omega}$ and an estimate of a possible proof (virtual proof) of L in $\mathbf{T}$. In order to produce this comparison we introduce a logical entropy measure H on proofs and axioms: H, which reflects the Shannon entropy measure (see [17,24]), expresses the qualitative aspects of the logical information included in a formal proof. The probability $p(L)$ is computed as $p(L) = H(Q)/H(P)$, Q virtual proof of L in $\mathbf{T}$, P suitable real proof of L in $\mathbf{\Omega}$. Q is obtained via the definition of an estimate criterion CR_T for the proofs in the system $\mathbf{T}$. The estimate criterion for proofs in $\mathbf{T}$ requires suitable proof-theoretic properties both of $\mathbf{T}$ and $\mathbf{\Omega}$, which are expressed by the notion of *regular logical calculus* $\mathbf{RK}$ and *regular theory* $\mathbf{T} = \mathbf{RK} + \mathbf{AxT}$, $\mathbf{AxT}$ proper axiom set. An *informational context* is therefore a setting $(\mathbf{\Omega}, \mathbf{AxT}, \mathbf{RK}, CR_T, H)$ where informational theorems $(L, p(L))$ are produced. We refer to [13–16], for the results already obtained as to classical settings. We limit ourselves to underline that in classical informational contexts $\mathbf{\Omega}$ and $\mathbf{T}$ are mutually consistent, and then the standard conjectures have a conservative character with respect to $\mathbf{T}$.

Here, the above outlined ideas will be discussed with reference to paraconsistent settings. In particular, we provide the proof-theoretical results that allow the extension of informational logic to a family of paraconsistent systems essentially given by the **C**-systems. The **C**-systems (see Carnielli–Marcos [7–9]) are paraconsistent systems allowing to formalize consistency and inconsistency statements by introducing a new monadic connective $\circ$, so that $\circ B$ has the intended meaning "B is consistent". We prove that suitable predicate logic sequent formulations of the **C**-systems are possible, that have the regularity properties required to support an informational context, thus allowing to obtain **C**-*informational contexts* (Definition 18) where conjectures on formal consistency and inconsistency can be generated. Moreover, since in a predicate logic setting the Provability Logic predicates $Pr_{\mathbf{V}}$ referring to any recursively axiomatized system $\mathbf{V}$ can be introduced, we may define contexts in which conjectures on consistency both from a local standpoint (i.e. centred on specific formulas) and from a global standpoint (i.e. referring to a theory) are simultaneously expressible.

Some external affinities between informational logic and uncertainty logics, possibilistic logics, fuzzy logic (e.g., Dubois [11], Kyburg [25]), formal logical investigations of probability (e.g., Halpern [23], Montagna [28], Scott–Krauss [33]) may be noted. However, we have to point out the deep difference with respect to our approach. The probability measure presented in Information Logic does not have a semantic foundation and does not reflect any uncertainty or vagueness of the epistemic state of the rational agent or of the given knowledge basis. Informational theorems $(B, p(B))$ are such that $p(B)$ is a probability of the *provability* of B in a given system $\mathbf{T}$, obtained through proof-theoretically based estimate criteria and logical information measures on the syntax of the real and virtual proofs on which we work. Thus, the informational theorem or formal conjecture $(B, p(B))$ is closer to an active proof process than to a vague knowledge assumption. A connection between the notion of probability of a proposition and its provability is proposed by Pearl in [30]; however, such approach is different since the notion of provability employed there has not a proof-theoretical foundation. As to the foundations and developments of the Paraconsistent Logic framework we refer to Batens et al. [1], Carnielli et al. [6], Da Costa [10]. As to the formalization of consistency and inconsistency statements through the **C**-systems we refer to the founding works of Carnielli–Marcos [5,7–9]. As to the formalization of global consistency and inconsistency statements about theories we refer to the work on Provability Logic by Boolos [2], Gentilini [18–20], Smorynski [34], Solovay [35]. We note that [18] is a seminal work with respect to the notion of formal conjecture of Informational Logic, since in it a mathematical notion of distance between a sequent and a system is introduced.

2. The Gentzen formulation of predicate logic systems

In this section we recall the sequent version **LK** of the classical predicate calculus and some sequent formulated paraconsistent systems existing in the literature. Our references for the general properties of the sequent calculi will be [4,18,21,36,37]. Accordingly, a sequent S is an expression of the form $X \vdash Y$ where X and Y are finite (possibly empty) sets of formulas. We will use the symbols $X, Y, \Lambda, \Gamma, \ldots$ as meta-expressions for sets of formulas, $A, B, C, D, \ldots$ for formulas. The intended meaning of a sequent $A_1, A_2, \ldots, A_n \vdash B_1, B_2, \ldots, B_m$ is $\bigwedge_i A_i \to \bigvee_j B_j$ and such equivalence holds both in a classical and in a paraconsistent setting. Given a rule $\frac{S_1 \ldots S_n}{S}$, the sequents $S_1, \ldots, S_n$ are the *premises* of the rule, the sequent S is the *conclusion* of the rule. The proofs are trees, whose leaves are axioms and whose branches are formed by sequent rules. In a proof-tree P a *branch* is a maximal linearly ordered set of sequents in it, having an axiom as the first element, where each sequent is a premise of the successive sequent. We also use the writings $\wedge X$, $\vee Y$ to indicate the conjunction (respectively disjunction) of the elements of X (respectively Y). We call the formula $\wedge X \to \vee Y$ *positive translation* of the sequent $X \vdash Y$. The writing Λ, Γ stands for $\Lambda \cup \Gamma$. If $X \subset \Lambda$ and $Y \subset \Gamma$ we say that $X \vdash Y$ is a *sub-sequent* of $\Lambda \vdash \Gamma$. The size $size(S)$ of a sequent S is the number of symbol occurrences in it. As to the definition of the set of subformulas of a given formula B, we specify that a strictly syntactic definition is used here for the subformulas of a quantified formula; that is, for example, if we have $\exists x \forall y A(x, y)$, A predicate letter, the subformula set is $\{A(x, y), \forall y A(x, y), \exists x \forall y A(x, y)\}$; moreover, subformulas of a given B that differ only by uniform renaming of bound variables are identified. In this exposition, for the sake of brevity, we will examine predicate calculi without the axioms for the equality predicate $= (.,.)$.

A theory **T** based on a sequent calculus **W** is given by the deduction apparatus of **W** plus a (possibly empty) proper axiom set **AxT** expressed by sequents. We also write **W** + **AxT** for **T**. The proofs of **T** are trees of sequents; the theorems of **T** are the roots of the trees. We also say that a formula A is a theorem of **T** if the sequent $\vdash A$ is a theorem of **T**. A theory **T** is *trivial* if it proves each sequent of the form $\vdash A$. **T** is *paraconsistent* if any formula B exists such that **T** plus $\vdash B \wedge \neg B$ is a non-trivial theory. We say that a set **U** of sequents *trivializes* **T**, and that **T** is *trivializable* by **U**, if **T** + **U** results as a trivial theory.

The sequent version **LK** of the classical predicate calculus is as follows:

Axioms: $A \vdash A$

Positive propositional logical rules:

$$\frac{B,\Gamma\vdash\Delta}{A\wedge B,\Gamma\vdash\Delta}\wedge-L \qquad \frac{B,\Gamma\vdash\Delta}{B\wedge A,\Gamma\vdash\Delta}\wedge-L \qquad \frac{\Gamma\vdash\Delta,A \qquad \Lambda\vdash X,B}{\Gamma,\Lambda\vdash\Delta,X,A\wedge B}\wedge-R$$

$$\frac{\Gamma\vdash\Delta,A}{\Gamma\vdash\Delta,A\vee B}\vee-R \qquad \frac{\Gamma\vdash\Delta,A}{\Gamma\vdash\Delta,B\vee A}\vee-R \qquad \frac{A,\Gamma\vdash\Delta \qquad B,\Lambda\vdash X}{A\vee B,\Gamma,\Lambda\vdash\Delta,X}\vee-L$$

$$\frac{A,\Gamma\vdash\Delta,B}{\Gamma\vdash\Delta,A\to B}\longrightarrow-R \qquad \frac{\Gamma\vdash\Delta,A \qquad B,\Lambda\vdash X}{A\to B,\Gamma,\Lambda\vdash\Delta,X}\longrightarrow-L$$

Negation rules: $$\frac{\Gamma\vdash\Delta,A}{\neg A,\Gamma\vdash\Delta}\neg-L \qquad \frac{A,\Gamma\vdash\Delta}{\Gamma\vdash\Delta,\neg A}\neg-R$$

Quantifier rules: $$\frac{[t/x]A,\Gamma\vdash\Delta}{\forall xA,\Gamma\vdash\Delta}\forall-L \qquad \frac{\Gamma\vdash\Delta,[b/x]A}{\Gamma\vdash\Delta,\forall xA}\forall-R$$

$$\frac{[b/x]A,\Gamma\vdash\Delta}{\exists xA,\Gamma\vdash\Delta}\exists-L \qquad \frac{\Gamma\vdash\Delta,[t/x]A}{\Gamma\vdash\Delta,\exists xA}\exists-R$$

where t is an arbitrary term and b is a free variable which does not occur in Γ, Δ. Moreover, t may be not fully quantified while b must be uniformly replaced by x (see [36]).

Structural rules:

Weakening rules: $$\frac{\Gamma\vdash\Delta}{\Gamma\vdash\Delta,A}W-R \qquad \frac{\Gamma\vdash\Delta}{A,\Gamma\vdash\Delta}W-L$$

Cut rule: $$\frac{\Gamma\vdash\Delta,A \qquad A,\Lambda\vdash X}{\Gamma,\Lambda\vdash\Delta,X}Cut$$

It is known that proofs of **LK** admit of *cut-elimination* (see [21,36,37]), i.e. each **LK**-proof P can be effectively transformed into a **LK**-proof Q without cut-rule occurrences having the same end-sequent.

3. Paraconsistent informational contexts

On one hand, informational contexts require estimate criteria for proofs of a sequent formulated system **V**, which are based on the proof-theoretical properties of **V**. On the other hand, the study of the sizes of the possible proofs of a given sequent or formula is presently a canonical topic in Proof Theory (see, e.g., [31]). Results and examples that can produce estimate criteria for proofs can be found in [29,31]. Results on estimate criteria for proofs in the informational logic framework are given in [13,16]. Our formal notion of estimate criterion for a sequent logical calculus is the following:

Definition 1. We say that a sequent formulated predicate calculus **W** having the same structural rules as **LK** admits of the *estimate criterion CR* for proofs if, fixed any language with at most a finite set of function symbols and individual constants, given a sequent L, given a bound k for the number of symbol occurrences in the terms (*complexity* of the terms), given a bound η for the number of formula occurrences in a sequent (*width* of the sequents), given a fixed finite sets of variables $V \equiv \{b_1,\ldots,b_m\}$,[1] *CR* is an effective procedure which produces, from the input (L,k,η,m):

[1] We canonically assume that the variables of the **W**-language are listed through a fixed sequence so that the number m univocally individuates the set $\{b_1,\ldots,b_m\}$.

(i) An estimate of the axiom set of a possible **W**-proof Q of L which respects the bounds k, η and m, that is a finite set $\{S_j\}$ of **W**-axioms such that, if some cut-free **W**-proofs of L within the bounds k, η and m exist, the axiom set of at least one of them is a subset of $\{S_j\}$;

(ii) An estimate of the set of the logical rule instances of a possible **W**-proof Q of L which respects the bounds k, η and m, that is a finite set $\{I_r\}$ of **W**-logical rules instances such that, if some cut-free **W**-proofs of L within the bounds k, η and m exist, the logical rule set of at least one of them is a subset of $\{I_r\}$;

(iii) An estimate of the length $l(Q)$ of a possible **W**-proof Q of L which respects the bounds k, η and m, that is an upper bound λ for the highest number of proof lines in a branch such that, if some cut-free **W**-proofs of L within the bounds k, η and m exist, at least one of them has each branch shorter than λ.

We call k, η and m the *parameters* of a *CR*-estimate.

Therefore, the estimate criterion *CR* is an effective procedure that, given L, k, η and m produces the mentioned estimate of a possible **W**-proof Q of L. Fixed *CR* and the input (L, k, η, m), it exists an unique estimated Q. Note that we do not impose that a real **W**-proof of L must exist. We remark that, in general, if **W** would include also the axiomatization of the equality predicate $=(.,.)$, then the condition "cut-free proof" must be changed into "proof containing at most atomic cuts" and this is the case of the calculus **EQ** in [13,16].

In general, fixed **W**, many estimate criteria may be definable. To discharge those *CR*'s which produce redundant estimated sets of values, we introduce the following notion, that allows in principle a mathematical distinction between significant and not relevant criteria:

Definition 2. Let **W** be a sequent predicate calculus as in Definition 1 above, where any language with at most a finite set of function letters and individual constants has been fixed. Let L be a **W**-provable sequent and let $\{P_i(L)\}$ be the set of the **W**-proofs of L. Let $\{S^i_j(L)\}, \{I^i_r(L)\}, \{\lambda^i(L)\}$ be respectively the set of $P_i(L)$-axiom set, of $P_i(L)$-rule set, of $P_i(L)$-length upper bound. Let $s_W(L) = \min_i\{card\, S^i_j(L)\}$, $v_W(L) = \min_i\{card\, I^i_j(L)\}$, $\lambda_W(L) = \min_i\{\lambda^i(L)\}$. Let $\{CR^t\}$ be the set of the estimate criteria for proofs in **W** and let $\{S^t_d(L,k,\eta,m)\}, \{I^t_z(L,k,\eta,m)\}, \{\lambda^t(L,k,\eta,m)\}$ be the sets of the corresponding estimates, fixed L and the estimate parameters k, η and m. Let $s_t(L,k,\eta,m) = \{card\, S^t_d(L,k,\eta,m)\}$ and $v_t(L,k,\eta,m) = \{card\, I^t_z(L,k,\eta,m)\}$. Then, a fixed CR^q is called a *most efficient estimate criterion for* **W** if, for each **W**-provable L and for each fixed k, η, m:

$$\begin{aligned}
&\Big\| \big(\big|s_q(L,k,\eta,m) - s_W(L)\big|, \big|v_q(L,k,\eta,m) - v_W(L)\big|, \big|\lambda^q(L,k,\eta,m) - \lambda_W(L)\big| \big) \Big\| \\
&\quad = \min_t \Big\{ \big\| \big(\big|s_t(L,k,\eta,m) - s_W(L)\big|, \big|v_t(L,k,\eta,m) - v_W(L)\big|, \\
&\qquad\qquad \big|\lambda^t(L,k,\eta,m) - \lambda_W(L)\big| \big) \big\| \Big\},
\end{aligned}$$

where $\|\,.\,\|$ is the euclidean norm.

Intuitively, a most efficient criterion is such that, for the **W**-provable sequents, it produces an axiom set, a rule set, and a length bound value that are in some sense minimal, compatibly with its effectiveness (in general recursive procedures such that, for each L, exactly the values $s_W(L)$, $v_W(L)$, $\lambda_W(L)$ are produced, may not exist, as in Section 3.1 is discussed). The notion suggests a strategy for a proof-theoretic research that has to select significant concrete criteria.

About the proposed concept of estimate criterion at least two further points must be discussed: (a) the dependence of the estimated proof on the parameters k, η, m; (b) the fact that, in the informational logic framework, the *CR*-estimate is relevant even if the fixed sequent L is not **W**-provable. We will examine such items in Section 3.1, after the presentation of the essential notions of the informational logic framework.

Definition 3. A *regular paraconsistent predicate calculus* **RPK** is a sequent predicate calculus such that:

(i) **RPK** is paraconsistent (see Section 2);
(ii) Axioms of **RPK** are the same as in the classical calculus **LK**;
(iii) Rules for the positive logical connectives and the structural rules of **RPK** belong to the set of the positive and structural rules of **LK**;
(iv) The set of the **RPK**-logical rules introducing the negation connective $\neg$ is not empty;
(v) **RPK** admits of cut-elimination and of an estimate criterion *CR* for proofs.

Note that the negation rule set of a regular paraconsistent **RPK** cannot include the classical one.

We also need estimate criteria for proofs in a theory **T** based on **RPK**. Then, it is necessary to introduce a notion of *normal proof* in a **RPK**-theory **T**, and this cannot be founded on cut-elimination. In a number of relevant theories based on a predicate calculus **W**, cuts cannot be eliminated: for example, the infinite arithmetical theories ranging between Primitive Recursive Arithmetic **PRA** and Peano Arithmetic **PA** do not admit of cut-elimination (see, e.g., [4,21]). However, for a number of them, a normal form for proofs is definable and a normal form theorem can be given, as e.g. is shown in [16]. The normality condition we adopt will be based on the distribution of the rules in the tree:

Definition 4. Let $\mathbf{T} \equiv \mathbf{RPK} + \mathbf{AxT}$ be a paraconsistent theory (see Section 2) based on a regular paraconsistent calculus **RPK**. Let the sequent $L \equiv X \vdash Y$ be **T**-provable through a proof P where the proper axiom instances $N_1, N_2, \ldots, N_s$ occur. Then, we say that P *can be reduced into a normal form* R if:

(i) a cut-free proof U of $B_1, B_2, \ldots, B_s, X \vdash Y$ in **RPK** exists, where the formulas B_j are the positive translations (see Section 2), possibly universally closed, of the sequents N_j;
(ii) R has U as the uppermost segment followed by a lower segment where each B_j in the U-root is deleted by a cut.

We say that R is a *normal* **T**-proof of L.

Definition 5. Let $\mathbf{T} \equiv \mathbf{RPK} + \mathbf{AxT}$ be a paraconsistent theory having a language with at most a finite set of function letters and individual constants and such that each **T**-proof admits of a normal form. Then, we say that **T** *admits of the estimate criterion* CR_T *for proofs* if CR_T is an effective procedure defined as *CR* in Definition 1, with the following specifications:

(a) "cut free **W**-proof" is replaced by "normal **T**-proof";
(b) as to the axiom set of a possible **T**-proof Q of a given sequent L, CR_T estimates the set of the proper **T**-axioms occurring in Q, and it does not consider the possible **RPK** axioms occurring in Q.[2]

Given the set $\{CR_T^u\}$ of the estimate criteria for proofs in **T**, a most efficient criterion CR_T^z is defined as in Definition 2. We note that the calculus **RPK** may admit of an estimate criterion *CR* and, however, a **RPK**-based theory **T** may exist that does not have any estimate criterion for proofs.

From an informational standpoint the proper axiom set $\mathbf{AxQ} \subset \mathbf{AxT}$ of a **T**-proof Q of a sequent L which is not a **RPK**-theorem plays a very important role, since it represents the specific information that is added to **RPK** in order to obtain L. Therefore, we refute the estimate criteria that produce a trivial estimate of the proper axiom set, that is:

Definition 6. Let CR_T be an estimate criterion for proofs of a theory $\mathbf{T} \equiv \mathbf{RPK} + \mathbf{AxT}$. Consider any input (L, k, η, m) of the procedure CR_T, such that:

(a) k, η, m are greater than $size(L)$;
(b) At least one formula B respecting the bounds k, m exists such that B occurs as formula or subformula in an instance of an element of **AxT** and B cannot be obtained by uniform term replacement from any formula or subformula in L.

Then, we say that CR_T is *acceptable* if the estimated proper axiom set $\{N_j\}$ is not the closure in the **T**-language of the sequent set **AxT** under the parameters k, η, m; CR_T *is not acceptable* otherwise.

Conditions (a) and (b) are imposed in order to avoid the trivial cases (e.g., the case where L is exactly the **AxT**-closure under k, η, m of **AxT**, or L has in the succedent the conjunction of all the formulas of the **AxT**-closure under k, η, m, and so on). For some classes of theories, the properties of the **T**-language and of the **T**-axioms may allow estimate criteria CR_T whose output does not depend on the parameters k, η, m:

[2] We observe that, as to the estimate criteria, a difference is established between the empty theory $\mathbf{RPK} + \emptyset$ and the calculus **RPK**, since the estimated axiom set will be always empty in the first case. Indeed, from the standpoint of informational logic, the proofs of the empty theory have an information input equal to 0, since a theory is characterized by the specific information added to the basic calculus.

Definition 7. We say that a criterion CR_T is *stable* if, for each sequent L, the CR_T-estimates do not depend on the parameters k, η, m. CR_T is *essentially stable* if the CR_T-estimated proper axiom set does not depend on the parameters k, η, m.

We shall see in Section 4 that at least one essentially stable CR_T exists.

Definition 8. A theory $\mathbf{T} \equiv \mathbf{RPK} + \mathbf{AxT}$ having a language with at most a finite set of function letters and individual constants is a *regular paraconsistent theory* if:

(i) **RPK** is a regular paraconsistent calculus, **AxT** is a recursive set and **T** is paraconsistent;
(ii) The **T**-proofs can be reduced to a normal form and **T** admits of an acceptable estimate criterion CR_T for proofs.

In Section 4 is explicitly presented (Definition 17) the acceptable estimate criterion for the proofs in theories **T** that are based on the regular paraconsistent calculus **BC**. Acceptable estimate criteria for proofs of theories based on the equational predicate calculus **EQ** are presented in [13,16].

Definition 9 (*Logical entropy measure of a formula*). Given a regular paraconsistent predicate calculus **RPK**, the *quantity of information* or *logical entropy* H of formulas and sequents is so defined:

(i) If A is an atomic formula of **RPK**, in which terms $t_1, \ldots, t_n$ occur, then:

$$H(A) = -\sum_i p_i \log p_i, \quad \text{where:}$$

$p_i = p[t_i] =$ (number of occurrences of t_i in A)/(number of occurrences of terms in A).

(ii) If B is a compound formula of **RPK** and $\{B_1, \ldots, B_m\}$ is the set of subformulas of B, where subformulas which differ only by uniform renaming of variables are identified, then:

$$H(B) = -\sum_j p[B_j] \log p[B_j] + \sum_k H(A_k), \quad \text{where:}$$

– $p[Bj] =$ (number of occurrences of B_j in B)/(number of occurrences of subformulas in B),
– $\{A_k\}$ is the set of the atomic formulas occurring in B.

(iii) If S is the sequent $A_1, A_2, \ldots, A_n \vdash B_1, B_2, \ldots, B_m$, then:

$$H(S) = \sum_j H(A_j) + \sum_i H(B_i);$$

(iv) If $\mathbf{B} \equiv \{B_1, \ldots, B_m\}$ is a set of formulas, then $H(\mathbf{B}) = \sum_i H(B_i)$;
(v) if $\mathbf{S} \equiv \{S_1, \ldots, S_n\}$ is a set of sequents, then $H(\mathbf{S}) = \sum_i H(S_i)$.

We wish to emphasize that the measure H cannot be replaced by more usual measures of complexity, of depth, and so on. Let us consider the following example:

Example 1. Let $A \equiv f(x) = f(y)$ and $B \equiv f(f(x)) = f(f(x))$. In human knowledge it is self-evident that the information wealth of A is greater than that of B. Nevertheless, if we employ as parameter the complexity $k =$ *highest number of symbol occurrences in a term* or the depth $d =$ *highest length of a decomposition tree of a term*, we have: $k(A) < k(B)$ and $d(A) < d(B)$. Conversely, the entropy measure H gives: $H(A) = 1.386 > H(B) = 1.098$. That is: a formula A may have greater entropy than that of a formula B with greater complexity and greater depth.

Definition 10 (*Logical entropy measure of a proof*). Let P be a normal proof of a sequent L in $\mathbf{RPK} + \mathbf{AxT}$, **RPK** regular paraconsistent calculus. The *logical entropy* $H(P)$ of the proof P is:

$$H(P) = \sum_i H(A_i) + \sum_r H(I_r),$$

$$H(I_r) = \left|H(\text{premises of } I_r) - H(\text{conclusion of } I_r)\right|, \quad \text{where:}$$

- $\{A_i\}$ is the set AxP of proper axiom instances occurring in P;
- $\{I_r\}$ is the set I_P of rule occurrences in P that introduce a logical symbol. We also write $H(I_P)$ for $\sum_r H(I_r)$.

We wish to comment briefly the entropy measures on formulas and proofs. First, we remark that it is an extension to a formal logical setting of the entropy measure of Shannon's Information Theory [17,24]. The entropy of a set of formulas measures, from a syntactic standpoint, the quality and the variety of the logical information enclosed, and disregards the merely quantitative or combinatorial aspects of the complexity of the formula. The example presented above is simple, but it illustrates clearly entropy as a quality oriented measure at the syntactic level. Moreover, the notion is founded on a parallel between axioms and proofs in Proof Theory and sources and channels in Information Theory. Thus, a set of formulas is always considered as the potential information input of a proof, and the entropy measure stresses this aspect. Following such parallel, the entropy measure of a proof has to emphasize the transformation power of the proof with respect to its information input. A wider discussion on the motivations of the syntactic entropy measure is presented in [13].

Definition 11. We call *paraconsistent informational context* a 5-tuple $(\mathbf{\Omega}, \mathbf{AxT}, \mathbf{RPK}, CR_T, H)$ such that:

(i) $\mathbf{T} \equiv \mathbf{RPK} + \mathbf{AxT}$ is a regular paraconsistent theory;
(ii) $\mathbf{\Omega}$ is a non-trivial undecidable system extending $\mathbf{T}$, such that the proof-theoretic strength (possibly expressed by ordinal measures: see [21,36,37]) of $\mathbf{\Omega}$ is greater than that of $\mathbf{T}$. Moreover, $\mathbf{\Omega}$ may be non classically consistent with $\mathbf{T}$;
(iii) CR_T is an acceptable estimate criterion for proofs in $\mathbf{T}$;
(iv) H is a logical entropy measure.

The most important mathematical notion which represents a rational intuition in the conjecture formulation process is that of virtual proof:

Definition 12 (*A priori virtual proof*). Let L be a sequent in the language of a regular paraconsistent theory $\mathbf{T} \equiv \mathbf{RPK} + \mathbf{AxT}$. Let CR_T be an acceptable estimate criterion for proofs in $\mathbf{T}$ and let k, η, m be any fixed estimate parameters. Then, we call *a priori virtual proof of L in* $\mathbf{T}$ (or *a priori virtual* $\mathbf{T}$*-proof of L*) the 4-tuple: $(AxQ, I_Q, H(Q), l(Q))$, where:

- AxQ is the set of proper axiom instances of the CR_T-estimated normal $\mathbf{T}$-proof of L;
- I_Q is the set of rules introducing logical symbols of the CR_T-estimated normal $\mathbf{T}$-proof of L;
- $H(Q)$ is $H(AxQ) + H(I_Q)$;
- $l(Q)$ is the length upper bound of the CR_T-estimated normal $\mathbf{T}$-proof of L.

We also write Q for the 4-tuple $(AxQ, I_Q, H(Q), l(Q))$ and $Q(k, \eta, m)$ to indicate explicitly the dependence on the parameters.

Virtual $\mathbf{T}$-proofs of L are independent of the existence of real $\mathbf{T}$-proofs of L. They are mathematical objects, since CR_T is an effective procedure representable by a recursive function. Moreover, since CR_T is an *acceptable* criterion, the virtual proofs cannot have a merely combinatorial nature. We remark that, fixed the criterion CR_T, given the parameters k, η, m, the virtual $\mathbf{T}$-proof of L is unique, even if many criteria may exist. Different criteria CR_T's give rise to different informational contexts based on the same $\mathbf{T}$, $\mathbf{RPK}$, H. Several questions arise about the notion of virtual proof. The main one is the following: what is the meaning of the virtual $\mathbf{T}$-proof of L in the cases in which L *is not* a $\mathbf{T}$-theorem? We discuss this point in Section 3.1.

Definition 13 (*Probability function for the* $\mathbf{T}$*-provability*). Let $\Pi \equiv (\mathbf{\Omega}, \mathbf{AxT}, \mathbf{RPK}, CR_T, H)$ be a paraconsistent informational context. Let L be a sequent in the language of $\mathbf{T}$, so that a real known proof P of L in $\mathbf{\Omega}$ exists. Let $[P]$ be the set of the $\mathbf{\Omega}$-proofs of L. Let $Q(k, \eta, m)$ be the above defined a priori virtual proof of L in $\mathbf{T}$. Let $H([P])$ be the value $\min_{R \in [P]}\{H(R)\}$. Then, we call *probability function for the* $\mathbf{T}$*-provability of L in the context Π* the function:

$$p(k, \eta, m)[L] = H\big(Q(k, \eta, m)\big)/H\big([P]\big).$$

Each time $H(Q(k, \eta, m)) > H([P])$, $p(k, \eta, m)[L]$ is normalized to 1.

We remark that each time CR_T is a stable criterion we have $p(k, \eta, m)[L] \equiv p[L]$, i.e. a unique probability value.

If the context Π is defined in order to produce applicative results in any automated deduction environment, it may be convenient to study the probability of L to be provable in $\mathbf{T}$ through the properties of the function $p(k, \eta, m)[L]$ of the estimate parameters, i.e. essentially as a measure depending on a sequence $\{Q_n\}$ of virtual proofs (see, for example, [15,16]). If the context Π is considered as the formalization of an epistemic environment and **T** is intended as an agent formulating conjectures, then we assume that proofs in **T** must have a constructive character with respect to the proofs in **Ω**. We suppose that such constructive feature is lost if the proof size exceeds some exponential bound γ with respect to the complexity of L and of the **T**-language, and then it is assumed that the estimated virtual proofs $Q(k, \eta, m)$ are not relevant if k, η, m are greater than γ:

Definition 14. Let $\Pi \equiv (\mathbf{\Omega}, \mathbf{AxT}, \mathbf{RPK}, CR_T, H)$ be a paraconsistent informational context and let $\xi \equiv$ number of function letters and individual constants in the **T**-language. Then, we call *constructive bound for virtual proofs of L in Π* the number: $\gamma(L, \mathbf{T}) \equiv (\mathit{size}(L) + \xi)!$

Definition 15 (*Informational theorems*). Let $\Pi \equiv (\mathbf{\Omega}, \mathbf{AxT}, \mathbf{RPK}, CR_T, H)$, L, P, $Q(k, \eta, m)$, $H([P])$ be as in Definition 13 above. Then, we call *probability of L to be provable in* **T** *in the context Π* the number:

$$p(L) = H(Q^*)/H\big([P]\big),$$

where $H(Q^*)$ is the mean value:

$$\frac{\sum_{\|(k,\eta,m)\|<\gamma(L,\mathbf{T})} H(Q(k, \eta, m))}{\mathit{card}\{(k, \eta, m)\colon \|(k, \eta, m)\| < \gamma(L, \mathbf{T})\}}$$

being $\|\,.\,\|$ the euclidean norm and $\gamma(L, \mathbf{T})$ the constructive bound for virtual proofs. We call $(L, p(L))$ *informational theorem* of the context Π.

We remark that if CR_T is a stable criterion then $p(k, \eta, m)[L] = p(L)$, i.e. a constant probability value.

In the following, we will refer to $(L, p(L))$ also with the expression "formal conjecture" with respect to **T**. We will also write briefly $p(L)$ for $p(k, \eta, m)[L]$ when the dependence on k, η, m is not relevant for the discourse.

3.1. Problems and discussions

(I) As to estimate criteria and estimate parameters, first, we wish to examine the dependence of estimate criteria *CR* and CR_T on the estimate parameters k, η, m that, in general, may be arbitrarily established. Indeed, for some interesting classes of theories, the normal form theorems for proofs can be so strong that provide *CR*- and CR_T-outputs which are essentially independent of the estimate parameters. However, let us consider the most general case in which the *CR*- or CR_T-output depends on the parameters k, η, m. We ask whether such dependence reflects the fact that the notion of estimate criterion is a weak notion, or if it expresses some essential limitative result of mathematical logic. We answer that it is a consequence of the undecidability of the full predicate logic and of many predicate logic based theories **T** having a deep mathematical meaning (**PRA**, **PA** and their subsystems, weak subsystems of second order Arithmetic, and so on, see [21,34,37]). As remarked in [3], we note that the parameters linked to the size of proofs "are important because they provide a measure of the difficulty of proving a given formula in a given formal system". But here is an important limitative result proven in [3]: *the λ-provability problem for a first* *order theory—i.e. given a formula A and an integer λ, to determine if A has a proof with λ or fewer lines—is undecidable*. Therefore, in general, we cannot have an effective method to establish that any size parameters k, η, m are the best in order to estimate the possible proof of a given sequent L. The approaches we have employed are mainly two:

(a) To consider k, η, m as essentially arbitrary and then to study the output of the procedure CR_T and the a priori virtual proof Q of L as functions of the parameters k, η, m; such approach produces the definition of probability function $p(k, \eta, m)[L]$ (Definition 13).
(b) To assume that the estimated proofs of L are not relevant if the estimate parameters k, η, m exceed some suitable constructive bounds, and then to consider only a suitable finite set of virtual proofs of L in the context $\Pi \equiv (\mathbf{\Omega}, \mathbf{AxT}, \mathbf{RPK}, CR_T, H)$. Such approach produces the definition of informational theorem $(L, p(L))$ of Π (Definition 15).

A crucial question could be the following: if the predicate calculus or the theory **T** admit cut-free proofs, why proof estimates independent of the size parameters k, η, m are in general not possible? Also the negative answer to such problem lies on the undecidability of the full predicate logic: the standard sequent calculi for predicate logic (the classical **LK**, the intuitionistic **LJ**, the linear **LL** [22], the paraconsistent $\mathbf{WG}_n$ [32]) admit of cut-elimination, but they are undecidable. A fundamental reason is that the classical subformula property of a cut-free proof is a poor tool in order to face the very difficult problem imposed by the fact that a cut-free proof P of a sequent L may contain *arbitrarily complex terms* which do not occur in the root L. This is due to the deletion power of the quantifier rules. For example, in a cut-free proof we may have the premise $\vdash A(t(r_1, \ldots, r_n))$ of a $\exists{-}R$ rule having $\vdash \exists x A(x)$ as the conclusion; the arbitrarily complex term occurrence $t(r_1, \ldots, r_n)$ is deleted. Analogously, we may have, in a cut-free proof P of L, a sequent S of the form $\vdash A(r_1), \ldots, A(r_n)$, n arbitrarily high, that, through n $\exists{-}R$ rules, gives rise to a lower sequent of the form $\vdash \exists x A(x)$; in this way, the width n of the sequent S is deleted. Therefore: the parameters k, η depend on the term complexity problem for proofs, which has a central role in producing the undecidability of predicate logic. We must moreover emphasize that it is not possible to avoid complex terms in order to obtain theories having a real mathematical expressive power, and reflecting the mathematical practice (consider e.g. the Primitive Recursive Arithmetic **PRA** whose terms express *all* the primitive recursive functions and procedures).

(II) As to virtual proofs, a fundamental question is the following: what is the meaning of the *CR*- or CR_T-output in the case in which the given sequent L is not **RPK**-provable or not **T**-provable? Such question naturally involves also the notion of a priori virtual proof, that is: what is the meaning of the a priori virtual **T**-proof Q of L if L is not **T**-provable ?

We preliminarily observe that, even if a real **T**-proof of L does not exist, the a priori virtual **T**-proof Q of L in the context is a defined mathematical object. Then, a first answer is that we are interested in establishing a metric between a sequent L and a given system or theory, based on the logical information measures. That is, the non **T**-provable sequents do not have the same status with respect to **T**: they are characterized by the quantity of specific information that must be added to **T** in order to prove them. Thus, as a preliminary result, we are satisfied if the CR_T-estimated a priori virtual **T**-proof of L contains any information that, suitably employed, allows formal measures of the distance between L and **T**. And this happens by the comparison with the **Ω**-proofs of L through the notion of syntactic probability $p(L)$ in the informational context. Q produces through $p(L)$ a contiguity or pertinence measure of L with respect to **T**. A formal notion of distance between a sequent and a system has been already introduced in [18] in order to prove the arithmetical completeness of modal logic. Moreover, from a different perspective, we may note that the a priori virtual proof Q also acquires a semantic role, in the sense of a constructive semantic a la Brower–Heyting: Q can be seen as a constructive meaning of L with respect to **T**, and $p(L)$ becomes a kind of intuitionistic truth value of L in the fixed context. However, the question above has been only postponed. We can formulated it again in the following form: is $p(L)$ a well-founded and efficient expression of the distance between L and **T**? And, moreover, is it a *useful* notion?

(III) Let us examine the theoretical and applicative meanings of Informational Contexts and probability functions. The question must be referred to the whole informational context $\Pi \equiv (\mathbf{\Omega}, \mathbf{AxT}, \mathbf{RPK}, CR_T, H)$ in which $p(L)$ is computed, either as probability function (Definition 13) or as probability value (Definition 15). The underlying philosophy of the informational context, both in the classical and in the paraconsistent cases, is the following: proofs in **Ω** of a given L are shorter and less constructive than that in **T**. Then, we suppose that the computational costs of the **Ω**-proofs of L are low, but the rules and axioms which are employed correspond to exacting mathematical and logical (and perhaps ontological) commitments. Conversely, we suppose that the **T**-proofs of L may be longer and with weaker axioms, corresponding to weaker mathematical (and perhaps epistemological, ontological) assumptions. The relationship between **Ω** and **T** in an informational context in some sense reproduces the one between classical and constructive mathematics. Many mathematical statements may not have complicated classical proofs (e.g., by the excluded middle principle) but it is not simple to establish if they have constructive proofs too, and possibly to write them. On the other hand, the proofs in **Ω** may be considered as "easy" in the sense that they may have a little size, but also as "complicated" in the sense that they may employ non-evident infinitary assumptions. After this, the reliability of $p(L)$ depends on a reasonable choice for **Ω** and **T**: **T** must admits of an efficient normal form theorem for proofs allowing a well-founded estimate criterion CR_T; a comparison theory for the proofs of **Ω** and **T** must be established, in order to obtain the measures $H(R)$ of a **T**-proof R and $H(P)$ of a **Ω**-proof P of L as homogeneous (and then, possibly, comparable) measures; minimality conditions for a **Ω**-proof of a given L must be defined. An example of such treatment for the Inductive Informational Contexts (i.e., briefly, contexts where **T** is an induction free theory and **Ω** is endowed by induction over **N**) is presented in [13,15,16].

If these requirements are satisfied, we can say that *the ratios* $H(Q(k,\eta,m))/H([P])$ or $H(Q^*)/H([P])$ *essentially express the quantity of information that* **T** *lacks in order to prove* L. The dependence on k,η,m in the functional expression tells us that our knowledge of this information gap will be meaningful only if we have at disposal a reliable estimate criterion. Each time we can conclude that such ratios are small (as to the functional expression this implies that the function $H(Q(k,\eta,m))$ is bounded) we can say that *the probability that* L *is* **T**-*provable is small*, and even that the distance (possibly we may think to such distance as $H([P]/H(Q(k,\eta,m)))$, $H([P]/H(Q^*)))$ between L and **T** is high. The name *probability* is justified by the fact that the selection of the criterion CR_T

and of the estimate parameters cannot be done through an effective (recursive) method. Having established the well foundedness of the notion of syntactic probability $p(L)$, we have to discuss its utility. We shall consider (i) a *theoretical* use and (ii) an *applicative* use.

(i) As to the theoretical use, we refer to the definition of informational theorem with $p(L) = H(Q^*)/H([P])$. We have already pointed out that it allows a notion of formal conjecture, in which the value $p(L)$ measures the reliability of the conjecture from the standpoint of the epistemic subject. For example, suppose that the following classical informational context has been defined: $\mathbf{\Omega}$ is Arithmetic **PA** plus transfinite induction $\mathbf{IND}(\varepsilon_0)$ on denumerable ordinals up to the ordinal ε_0 [37]; **T** is Robinson Arithmetic **RA**, i.e. a very weak induction free subsystem of **PA** having an important role in constructive mathematics. Then the context has the form $(\mathbf{PA} + \mathbf{IND}(\varepsilon_0), \mathbf{AxRA}, \mathbf{LK}, CR_{RA}, H)$. Let us identify the epistemic subject with the theory **RA**. Let us consider these possible conjectures of the subject: $Con(\mathbf{RA})$, $Con(\mathbf{PRA})$, $Con(\mathbf{PA})$.[3] **RA** does not prove any of them; *however they are very different conjectures from the standpoint of* **RA**. The first two ones imply weak inductions over **N**; the third one implies transfinite induction up to the ordinal ε_0. If the informational context is well founded, we expect that $Con(\mathbf{RA})$, $Con(\mathbf{PRA})$ result as closer to **RA** than $Con(\mathbf{PA})$, and that $Con(\mathbf{PA})$ results as very distant from **RA** (i.e., presumably, $p(Con(\mathbf{RA})) > p(Con(\mathbf{PRA})) \gg p(Con(\mathbf{PA}))$). Observe that it would be not simple to formalize the epistemological difference among the three statements through semantical tools: all three are true in (a suitable expansion of) the standard model of **RA**. Then, $Con(\mathbf{PRA})$ is a more reliable conjecture than $Con(\mathbf{PA})$ for the agent **T**, and in a possible reasoning on conjectures the different reliability grades must be taken into account.

Observe, moreover, that in this case the estimate criterion CR_{RA} and the computation of the a priori virtual **RA**-proofs must be applied to sentences which are not **RA**-provable. In fact, the example above allows us also to emphasize a fundamental property of the informational contexts: even if $\mathbf{T} \nvdash B$, the epistemic subject **T** cannot know that it does not prove B; indeed, by Gödel theorems, no **T** can prove $\neg Pr_{\mathbf{T}}(\ulcorner B \urcorner)$,[4] whatever B is. Therefore, for example, the above considerations on the non **RA**-provability of the sentences $Con(\mathbf{RA})$, $Con(\mathbf{PRA})$, $Con(\mathbf{PA})$ belong to the metatheory about the context, and cannot never be formulated by $\mathbf{T} \equiv \mathbf{RA}$. Then, all three sentences may be only proper conjectures from the standpoint of **T**, which may apply to them the effective estimate procedure at disposal.

(ii) Let us consider the applicative use of the syntactic probability measure. Such use is exemplified in [16] for the automated deduction in the inductive informational contexts, but the approach may be in principle generalizable. Indeed, the notions of syntactic probability and of informational context have been originated by the following problem: given a theory **T** and a sequent L, to find a non-standard (implementable) decision method in order to establish if L is **T**-provable, without necessarily proving L in **T**. It is assumed that the computational costs of the possible **T**-proof of L are high, while the automated deduction apparatus may efficiently produce proofs of L in a suitable **T**-extension $\mathbf{\Omega}$. In order to produce such decision theory the syntactic probability must be studied as a probability function $p(k, \eta, m)[L]$: considering the CR_T-output $CR_T(L, k, \eta, m)$ as a function of the

[3] We write $Con(\mathbf{V})$ for the formula expressing in the arithmetical formal language the consistency of the system **V**.

[4] Recall that $\neg Pr_{\mathbf{T}}(.)$ is the provability predicate for **T** and that $Pr_{\mathbf{T}}(\ulcorner B \urcorner)$ means "B is not **T**-provable" [34].

parameters k, η, m, a sequence $\{Q_n\}$ of a priori virtual **T**-proofs of L can be produced. Moreover, introducing a suitable notion of sequence expansion of a proof P in $\mathbf{\Omega}$ it is possible to generate a sequence $\{P_n\}$ of $\mathbf{\Omega}$-proofs of (suitable instances of) L. Thus, we can consider also the sequence p_n of the entropy ratios $H(Q_n)/H(P_n)$. The study of the properties of p_n and of other information measure sequences computed on Q_n, P_n, allows to obtain the searched decision criteria.

4. A proof-theoretical analysis of the C-systems

In this section we introduce a predicate sequent version **BC** of the *paraconsistent basic logic of formal inconsistency* **bC** defined in Carnielli–Marcos [7]. In a predicate logic environment we call *local* consistency or inconsistency sentence each sentence including formulas of the form $\circ B$, $\neg(A \wedge \neg A)$, $B \wedge \neg B$ as subformulas. The interest of the notion is also given by the fact that it may interact with a notion of *global* consistency or inconsistency statement, referring to any recursively axiomatized system **V**, through the provability predicate $Pr_{\mathbf{V}}(.)$ (see [2,18,34]). For example, in a predicate logic environment we can study either classical consistency statements of the form $\neg Pr_{\mathbf{V}}(\ulcorner A \wedge \neg A \urcorner)$ or formulas $Pr_{\mathbf{V}+B}(\ulcorner A \wedge \neg A \urcorner) \wedge \exists x \neg Pr_{\mathbf{V}+B}(x)$ asserting the paraconsistency of **V** and of the theory $\mathbf{V} + B$. Thus, predicate logic theories of the form $\mathbf{BC} + \mathbf{AxT}$ can be the most expressive setting for the formalization of meta-reasoning about consistency and inconsistency, both in its global and its local aspects. Moreover, we will show that **BC** is a regular paraconsistent predicate calculus and that the predicate logic systems corresponding to the **C**-systems presented in [7] can be seen as regular paraconsistent theories. Thus, the local consistency or inconsistency statements studied by the **C**-systems extending **bC**, *can be studied also as formal conjectures* in the paraconsistent informational contexts based on **BC**. To this purpose, we must carry out a cut-elimination theorem for **BC**, and a set of proof-theoretical results that allow to establish the desired regularity properties for a relevant class of **BC**-based theories.

The calculus **BC** is defined as follows: axioms, positive propositional logical rules, quantifier rules, structural rules, are the same as in the calculus **LK** presented in Section 2; the negation rules are the followings:

$$\frac{A, \Gamma \vdash \Delta}{\neg\neg A, \Gamma \vdash \Delta}\,\neg{-}L1 \qquad \frac{\circ A, \Gamma \vdash \Delta, A}{\circ A, \neg A, \Gamma \vdash \Delta} \longrightarrow \neg{-}L3 \qquad \frac{A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg A}\,\neg{-}R.$$

We call the formula $\circ A$ in the rule $\neg{-}L3$ *constraint formula* of the rule. It is well known that in **bC** and in the **C**-systems the connective $\circ$, having the intended meaning "A is consistent", is not definable starting from the other logical connectives. Moreover, we remark that in **BC** the axioms $A \vdash A$ cannot be restricted to the atomic case only. For example, due to the constraints in $\neg{-}L3$, we cannot prove $\neg A \vdash \neg A$ from $A \vdash A$ in **BC**.

Proposition 3. *The system* **BC** *has exactly the same theorems of* **bC**, *provided that theorems $X \vdash Y$ of* **BC** *are translated into formulas $\wedge X \to \vee Y$ of the* **bC**-*language, and that a suitable standard translation of* **bC**-*formulas into sequents is assumed.*

Proof. From [7] we know that **bC** is given by the minimal paraconsistent system $\mathbf{C}_{\min}$ plus the Gentle Principle of Explosion **bc1**: from $\circ A \wedge (A \wedge \neg A)$ each formula B is provable. It is obvious that **BC** proves the $\mathbf{C}_{\min}$ theorems. We establish that a sequent version of the principle **bc1** is: $\circ A, A, \neg A \vdash$. Then, the following is the **BC**-proof of **bc1**:

$$\cfrac{\cfrac{A \vdash A}{\circ A, A, \vdash A}}{\circ A, A, \neg A \vdash}$$

It is known that **BC** *minus* $\neg{-}L3$ is a sequent version of $\mathbf{C}_{\min}$ [32]. Then, we have to show that **BC** *minus* $\neg{-}L3$ plus (**bc1**) proves each conclusion of a $\neg{-}L3$ rule. Indeed, let $\frac{\circ A, \Gamma \vdash \Delta, A}{\circ A, \neg A, \Gamma \vdash \Delta}$ be any $\neg{-}L3$ rule occurrence in a **BC**-proof; then, it can be replaced by the following cut:

$$\frac{\circ A, \Gamma \vdash \Delta, A \qquad \circ A, A, \neg A \vdash}{\circ A, \neg A, \Gamma \vdash \Delta} \qquad \square$$

In order to expose the cut-elimination theorem for **BC** we need to recall some notions of proof-theory:[5]

Definition 16. (i) In a proof-tree P in **BC** the *depth* or *height* $h(P)$ is the highest number of proof-lines in a branch. The *grade* $g(A)$ of a formula A is the number of occurrences of logical symbols in it.

(ii) In a rule occurrence R in a proof-tree P in **BC** we call: *auxiliary formulas* the formula occurrences in the premises on which the rule acts; *principal formula*, or *formula introduced by the rule*, the formula occurrence produced by the rule in the conclusion. Each formula in the conclusion of R is called the *successor* of the formulas in the premises corresponding to it, that are called its *predecessors*. In a branch of P we say that the formula occurrence B is an *ancestor* of the formula C occurring below B in the branch, called a *descendant* of B, if they are connected by a sequence of predecessor–successor relations alongside the branch. C is called an *integral descendant* of B if B and C are the same formula.

Theorem 1. *Cut elimination holds for* **BC**.

The proof is shown in Appendix A.

At this point a question arises: does a sequent formulated predicate logic extension of the **C**-system **Ci** (see [7]) exist, having the same regularity properties as **BC**? We believe that the answer is negative, since it seems to be very difficult to define cut-free sequent formulations of **Ci** which are not redundant, i.e. that do not include a lot of *ad hoc* rules. A sequent formulation of **Ci** is **BC** plus the axiom $\neg\circ A \vdash A \wedge \neg A$. It is evident that such version cannot admit cut-elimination. But we also have the following general limitative result:

[5] Our definitions are similar but not identical to that of Troelstra–Schwichtenberg [37] and, subordinately, to that of Takeuti [36].

Proposition 4. *It is not possible to define a cut free sequent formulation* **H** *of* **Ci** *by only changing the set of the rules of* **BC** *that introduce the negation connective* $\neg$.

Proof. Let us suppose *ad absurdum* that **H** exists. Then it must be consistent and has the sequent $\vdash^{\circ\circ} A$ as the root of a cut-free proof. But $^{\circ\circ}A$ in the root may be neither the principal formula of a weakening with the empty sequent as premise, nor the principal formula of any arbitrary negation rule, and in both cases we have an absurd. □

Therefore, we prefer to consider the system **Ci** given by **BC** plus $\neg^{\circ}A \vdash A \wedge \neg A$ as a regular paraconsistent theory based on **BC**. Then, it must be concluded the proof that **BC** is a regular paraconsistent predicate calculus that can support regular theories.

Proposition 5. BC *is a regular paraconsistent predicate calculus.*

Proof. The proof is essentially identical to that given in [16] for the classical calculus **LK**. □

Conversely, in order to characterize the regular paraconsistent theories based on **BC**, we must give specific proofs.

Lemma 1. *Let* $X \vdash Y$ *be a sequent in the* **BC**-*language. Then,* $\wedge X \to \vee Y, X \vdash Y$ *is* **BC**-*provable through positive propositional inferences only.*

The proof is straightforward.

Theorem 2. *Let* $\mathbf{T} \equiv \mathbf{BC} + \mathbf{AxT}$ *be a paraconsistent theory,* $\mathbf{AxT} \equiv \{X_i \vdash Y_i\}_{i \in I}$ *recursive proper axiom set, having a language with at most a finite set of function letters and individual constants. Then,* **T** *is regular.*

Proof. Without loss of generality, we assume that in **AxT** only sentences occur. Let $U \vdash V$ be the root of a **T**-proof P. Lemma 1 allows us to replace each proper axiom occurrence $X_j \vdash Y_j$ in P with the **BC**-proof having $\wedge X_j \to \vee Y_j$, $X_j \vdash Y_j$ as end sequent. Then, after possible suitable renaming of free variables in the branches and possible eliminations of cuts, we obtain a cut-free proof P^* in **BC** of the sequent $\{\wedge X_j \to \vee Y_J\}, U \vdash V$, being $\{X_j \vdash Y_i\}_{j \in J}$ the set of the proper **T**-axiom occurrences in P. Since the sequents $\vdash \wedge X_j \to \vee Y_j$ are **T**-theorems trivially provable from each proper axiom $X_j \vdash Y_j$, if we apply d suitable cuts starting from the root of P^*, $d = card(J)$, we obtain a **T**-proof P^{**} of $U \vdash V$ which is a normal **T**-proof, as defined in Definition 4. Moreover, by construction of the cut-free segment P^* of P^{**}, each formula in $\{X_j \vdash Y_j\}_{j \in J}$ can be obtained through a suitable term-replacement from a subformula of a formula occurring in $U \vdash V$. Then, given a sequent $W \vdash Z$, fixed the bounds k, η, and the variable set $\mathcal{V} \equiv \{b_1, \ldots b_m\}$ as mentioned in Definitions 1 and 5, we can estimate the set of proper **T**-axiom occurrences in a possible normal **T**-proof Q of $W \vdash Z$ as included in $\mathcal{A}$, where $\mathcal{A}$ is the set of *proper axiom instances* $\Delta_r \vdash \Pi_r$ such that: (i) they have formulas that can be obtained through a suitable term-replacement from a subformula of a formula occurring in $W \vdash Z$; (ii) they respect the bounds k, η; (iii) they include possible variables that range among $b_1, \ldots, b_m$. Let $n = card(\mathcal{A})$. Let $\mathcal{G} \equiv \{t_1, t_2, \ldots, t_s\}$ be the set of terms in the **T**-language, with complexity $\leqslant k$ and variables ranging in $\mathcal{V}$. Assuming that the possible **T**-proof Q of $W \vdash Z$ is normal, we have that the greatest length of a branch in Q is $\leqslant n + card(\mathcal{E})$, where $\mathcal{E}$ is the set of the sequents of width $\leqslant \eta$ including formulas obtained from subformulas of formulas occurring in $W \vdash Z$ by term-replacements with terms ranging among $t_1, t_2, \ldots, t_s$. $\mathcal{E}$ includes the sequents occurring in Q. The set of the logical rule occurrences in Q can be estimated as included in the set of the **BC**-logical rule instances having the principal formula that can be obtained through a suitable term-replacement from a subformula of a formula occurring in $W \vdash Z$, and their premises and their conclusion belonging to $\mathcal{E}$. □

The proof of Theorem 2 allows to define the following estimate criterion:

Definition 17. Let $\mathbf{T} \equiv \mathbf{BC} + \mathbf{AxT}$ be a paraconsistent theory, $\mathbf{AxT} \equiv \{X_i \vdash Y_i\}_{i \in I}$ recursive proper axiom set, having a language with at most a finite set of function letters and individual constants. Let $L \equiv W \vdash Z$ and let k, η, m be the fixed estimate parameters. Then, the estimate criterion CR_T is the following procedure, assuming the basic case where only sentences occur in **AxT**:

1. If $L \in \mathbf{AxT}$ or $L \equiv {\vdash}{\wedge}X$, such that formulas in X are the positive translations of sequents of a set $\mathbf{U} \subset \mathbf{AxT}$, then write L and stop; if $L \equiv {\vdash}{\vee}Y$, such that formulas in Y are the positive translations of sequents of a set $\mathbf{V} \subset \mathbf{AxT}$, write ${\vdash}A$, A any element of Y, and stop.
2. If $L \notin \mathbf{AxT}$ or L is different from each ${\vdash}{\wedge}X$, ${\vdash}{\vee}Y$ of the form described in point 1, then:
   2.1. Write the set $\mathcal{D} \equiv \{t_1, \ldots, t_s\}$ such that t_i is a term of the **T**-language containing variables at most occurring in $\{b_1, \ldots, b_m\}$ and with complexity $\leqslant k$;
   2.2. Write the set $\mathcal{F} \equiv \{A_u\}$, A_u formula of the **T**-language such that $terms(A_u) \subset \mathcal{D}$ and A_u is obtained by a uniform term replacement from a subformula of a formula occurring in $W \vdash Z$;
   2.3. Write the set $\{\Delta_r \vdash \Gamma_r\}$ such that $width(\Delta r \vdash \Gamma r) \leqslant \eta$, $\Delta r \vdash \Gamma r$ is an instance of an axiom of **AxT**, each subformula of a formula occurring in $\Delta r \vdash \Gamma r$ belongs to $\mathcal{F}$. $\{\Delta r \vdash \Gamma r\}$ is the proper axiom set output of CR_T;
   2.4. Write the set $\mathcal{E}_{\text{prem}} \equiv \{X_g \vdash Y_g\}$ such that $width(X_g \vdash Y_g) \leqslant \eta$, subformulas of formulas occurring in $X_g \vdash Y_g$ belong to $\mathcal{F}$, $X_g \vdash Y_g$ can be a premise of a logical rule of **BC**;
   2.5. Write the set $\mathcal{E}_{\text{con}} \equiv \{U_f \vdash V_f\}$ such that $width(U_f \vdash V_f) \leqslant \eta$, subformulas of formulas occurring in $U_f \vdash V_f$ belong to $\mathcal{F}$, $U_f \vdash V_f$ can be the conclusion of a logical rule of **BC**;
   2.6. Write the set $\{I_z\}$ such that I_z is a **BC**-logical rule instance with premises $\in \mathcal{E}_{\text{prem}}$ and the conclusion $\in \mathcal{E}_{\text{con}}$. $\{I_z\}$ is the logical rule set output of CR_T;
   2.7. Write the set $\mathcal{E} \equiv \{\Sigma_c \vdash \Theta_c\}$ such that $width(\Sigma_c \vdash \Theta_c) \leqslant \eta$, subformulas of formulas occurring in $\Sigma_c \vdash \Theta_c$ belong to $\mathcal{F}$. Write the number: $\lambda \equiv card\{\Delta_r \vdash \Gamma_r\} + card\,\mathcal{E}$. λ is the length upper bound output of CR_T.

The extension to the case where **AxT** includes open formulas is straightforward.

By recalling the proof of Theorem 2 we can prove that:

Proposition 6. *CR_T is an acceptable estimate criterion.*

We can also show that a relevant class of theories $\mathbf{T} \equiv \mathbf{BC} + \mathbf{AxT}$ having essential stable estimate criteria exists.

We say that a theory **T** is *purely predicative* if in the **T**-language function letters do not occur. Thus, it is possible to prove that:

Proposition 7. *Let $\mathbf{T} \equiv \mathbf{BC} + \mathbf{AxT}$ be a paraconsistent theory such that the **T**-language is purely predicative and with at most a finite number of individual constants, and such that in **AxT** only closed formulas occur. Then, **T** has an estimate criterion CR_T which is essentially stable.*

The interest of the result is due to the fact that, as it is known, many relevant theories can be reformulated in a purely predicative language:

Example 2. Let **AxT** be the Peano axioms for arithmetical systems in the language $\{0, S(.), +(.,.) \cdot (.,.), = (.,.)\}$. Then, we can replace each term $S(t)$ with $t + 1$ and then, as e.g., in [29] is proposed, the functions $+(.,.) \cdot (.,.)$ with two three-place predicates $A(.,.,.)$ and $B(.,.,.)$ respectively, such that, for example, any formula $x + y = z$ is translated into $A(x, y, z)$ and so on. **AxT** can be reformulated in the new language $\{0, 1, A(.,.,.), B(.,.,.), = (.,.)\}$.

Definition 18. We call *paraconsistent informational context for conjectures on formal consistency* a context of the form $(\mathbf{\Omega}, \mathbf{AxT}, \mathbf{BC}, CR_T, H)$ where **T** has the form $\mathbf{BC} + \neg{\circ}A \vdash A \wedge \neg A + \{X_i \vdash Y_i\}_{i \in I}$, I possibly empty and CR_T is the criterion defined in Definition 17. A **C**-*context* is a context of the form $(\mathbf{K}, \mathbf{AxCi}, \mathbf{BC}, CR_T, H)$ where **K** is any sequent formulation of the predicate logic version of any **C**-system extending **Ci** among those presented in [7] and **AxCi** is $\neg{\circ}A \vdash A \wedge \neg A$. A *basic* **C**-context has the form $(\mathbf{K}, \emptyset, \mathbf{BC}, CR_T, H)$ with **K** possibly identical to **Ci**.

The above defined context will be the basis for the introduction of a proof-theory acting on conjectures [12]. The paraconsistent informational contexts also acquire interesting epistemological corroborations. Starting from the consideration that inconsistencies occur in the construction processes of scientific theories (see, e.g., Meheus [26,27]), we show in [12] that the conjectural inference we propose may contribute to the logical characterization of the scientific discovery.

Appendix A

Theorem 1. *Cut elimination holds for BC.*

Proof. We consider an uppermost cut C in a proof-tree P in BC and show that it can be replaced by BC-deductions without cuts. The proof is by induction on the cutrank with a subinduction on the level of C, that are so defined: the cutrank of C is the grade of the cut formula in C, the level of C is the sum of the depths of the deductions of the premises. Let C be the following cut in P:

$$\dfrac{\begin{array}{c}Q1\\ \Gamma \vdash \Delta, A\end{array} \qquad \begin{array}{c}Q2\\ A, \Lambda \vdash X\end{array}}{\Gamma, \Lambda \vdash \Delta, X},$$

where $Q1$ and $Q2$ are the P-sub proofs of the premises. The only cases that we have to examine are the ones in which the cut formula A has either the form $\neg B$ or the form $\circ B$. As to the other cases, we refer to the cut elimination proofs for classical predicate calculus LK (see [21,36,37]). We previously establish the following preliminary reductions on the tree P: if one of the cut-formulas of C has the form $\neg B$ and is the descendant of the principal formula F of a weakening rule in P, we delete such weakening and each rule in P acting on a descendant of F, by getting a cut-free proof of a sub-sequent of $\Gamma, \Lambda \vdash \Delta, X$; analogously, if the left cut formula of C has the form $\circ B$ and is the descendant of the principal formula of a weakening rule, we delete such weakening, getting a cut-free proof of a sub-sequent of $\Gamma, \Lambda \vdash \Delta, X$. After such preliminary reductions, we consider these different cases:

(1) At least one of $Q1$, $Q2$ is an axiom. Let the left premise be an axiom of the form $A \vdash A$. Then, the conclusion of C has the form: $A, \Lambda \vdash X$. Then, we replace C by the right premise. If the right premise is an axiom, we replace the cut by the left premise.

(2) Neither $Q1$ nor $Q2$ is an axiom and the cut formula is not principal in at least one of the premises. If the cut formula is not on the both sides principal, let us consider for example the premise $Z \vdash W, A$ of the one-premise rule R in $Q1$ having the left cut premise $\Gamma \vdash \Delta, A$ as conclusion. Then we produce the following proof:

$$\dfrac{\dfrac{Z \vdash W, A \qquad A, \Lambda \vdash X}{Z, \Lambda \vdash W, X}R}{\Gamma, \Lambda \vdash \Delta, X},$$

where the level of the introduced cut is lower than that of C, and the cutrank is the same. If R is a two-premise rule the reduction is similar, and so it is for the sub-cases in which the right cut formula is not principal, with the exception of the case in which A has the form $\circ B$ and the right cut premise is the conclusion of a $\neg{-}L3$ rule K having $\circ B$ as constraint formula. In this case let us consider the sub-proof $Q1$ of the left C-premise $\Gamma \vdash \Delta, \circ B$ in P. We observe that, by hypotheses, the occurrence of $\circ B$ in such premise cannot be the principal formula of any rule in $Q1$ and then it must be introduced by a set of axiom occurrences of the form $\circ B \vdash \circ B$; moreover, it must be the integral descendant of the right formula of each axiom. If we replace each axiom occurrence by the proof of the C-premise $\circ B, \Lambda \vdash X$, we have a cut-free proof of $\Gamma, \Lambda \vdash \Delta, X$, after possible suitable renaming of free variables in the branches. We observe that no $(\neg{-}L3)$-rule constraints are broken in the reduction, since a $\circ B$ in the succedent cannot be a constraint formula in any $\neg{-}L3$.

(3) The cut formula is principal in both the premises of the cut C. If the cut formula has the form $\circ B$, then both the premises of the cut are the conclusions of weakening rules, and

this case has already been solved by the preliminary reductions. Let us consider the cases in which the cut formula has the form $\neg B$. By hypotheses, both occurrences of $\neg B$ are the principal formulas of negation rules.

(3.1) The left premise is the conclusion of a $\neg-R$ rule and the right premise is the conclusion of a $\neg-L1$ rule:

$$\frac{\dfrac{\neg B, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg\neg B} \qquad \dfrac{B, \Lambda \vdash X}{\neg\neg B, \Lambda \vdash X}}{\Gamma, \Lambda \vdash \Delta, X}$$

Then, we replace the cut C with the following proof:

$$\frac{\dfrac{B, \Lambda \vdash X}{\Lambda \vdash X, \neg B} \qquad \neg B, \Gamma \vdash \Delta}{\Gamma, \Lambda \vdash \Delta, X}$$

where the cutrank is lower.

(3.2) The left premise of C is the conclusion of a $\neg-R$ rule and the right premise is the conclusion of a $\neg-L3$ rule:

$$\frac{\dfrac{B, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg B} \qquad \dfrac{{}^{\circ}B, W \vdash X, B}{{}^{\circ}B, \neg B, W \vdash X}}{{}^{\circ}B, \Gamma, W \vdash \Delta, X}$$

Then we replace the cut C with the following proof:

$$\frac{{}^{\circ}B, W \vdash X, B \qquad B, \Gamma \vdash \Delta}{{}^{\circ}B, \Gamma, W \vdash \Delta, X}$$

where the cutrank is lower. □

References

[1] D. Batens, C. Mortensen, G. Priest, J.P. Van Bendegem (Eds.), Frontiers of Paraconsistent Logic, Proceedings of WCP I, Research Studies Press Limited, Baldock, UK, 2000.
[2] G. Boolos, The Logic of Provability, Cambridge University Press, 1993.
[3] S.R. Buss, The undecidability of k-provability, Ann. Pure Appl. Logic 53 (1991) 75–102.
[4] S.R. Buss, An introduction to proof theory, in: S.R. Buss (Ed.), Handbook of Proof Theory, Elsevier, Amsterdam, 1998, pp. 2–78.
[5] W.A. Carnielli, On sequents and tableaux for many-valued logics, J. Non-Classical Logic 8 (1) (1991) 59–78.
[6] W.A. Carnielli, M.E. Coniglio, I.M.L. D'Ottaviano (Eds.), Paraconsistency: the Logical Way to the Inconsistent, Proceedings of WCP 2000, Dekker, New York, 2002.
[7] W.A. Carnielli, J. Marcos, A taxonomy of C-systems, in: W.A. Carnielli, M.E. Coniglio, I.M.L. D'Ottaviano (Eds.), Paraconsistency: the Logical Way to the Inconsistent, Proceedings of WCP 2000, Dekker, New York, 2002, pp. 1–94.
[8] W.A. Carnielli, J. Marcos, Limits for paraconsistent calculi, Notre Dame J. Formal Logic 40 (3) (1999).
[9] W.A. Carnielli, J. Marcos, Tableau systems for logics of formal inconsistency, in: H.R. Arabnia (Ed.), Proceedings of the 2001 International Conference on Artificial Intelligence, vol. II, CSREA Press, USA, 2001, pp. 848–852.
[10] N.C.A. Da Costa, On the theory of inconsistent formal systems, Notre Dame J. Formal Logic XV (4) (1974) 497–510.
[11] D. Dubois, J. Lang, H. Prade, Possibilistic logic, in: M. Gabbay Dov, C.J. Hogger, J.A. Robinson (Eds.), Handbook of Logic in AI and Logic Programming, vol. 3, Clarendon Press, Oxford, 1994.
[12] P. Forcheri, P. Gentilini, Paraconsistent Conjectural Deduction based on Logical Entropy Measures (in preparation).

[13] P. Forcheri, P. Gentilini, M.T. Molfino, Research in automated deduction as a basis for a probabilistic proof-theory, in: P. Aglianò, A. Ursini (Eds.), Logic and Algebra, Dekker, New York, 1996.
[14] P. Forcheri, P. Gentilini, M.T. Molfino, Informational logic for automated reasoning, in: J.J. Alferes, L.M. Pereira, E. Orlowska (Eds.), Logics in Artificial Intelligence, Springer Verlag, 1996.
[15] P. Forcheri, P. Gentilini, M.T. Molfino, Informational logic as a tool for automated reasoning, J. Automated Reas. 20 (1998) 167–190.
[16] P. Forcheri, P. Gentilini, M.T. Molfino, Informational logic in knowledge representation and automated deduction, Artificial Intelligence Comm. 12 (1999) 185–208.
[17] R. Gallager, Information Theory and Reliable Communication, Wiley, New York, 1968.
[18] P. Gentilini, Proof-theoretic modal PA-completeness. I: A system-sequent metric, Studia Logica 63 (1999) 27–48.
[19] P. Gentilini, Proof-theoretic modal PA-completeness. II: The syntactic countermodel, Studia Logica 63 (1999) 245–268.
[20] P. Gentilini, Proof-theoretic modal PA-completeness. III: The syntactic proof, Studia Logica 63 (1999) 301–310.
[21] J.Y. Girard, Proof-Theory and Logical Complexity, Bibliopolis, Napoli, 1987.
[22] J.Y. Girard, Linear logic, TCS 50 (1987) 1–102.
[23] J.Y. Halpern, An analysis of first order logics of probability, Artificial Intelligence 46 (1990) 311–350.
[24] A.I. Khinchin, Mathematical Foundations of Information Theory, Dover Publications, New York, 1957.
[25] H.E. Kyburg, Uncertainty logics, in: M. Gabbay Dov, C.J. Hogger, J.A. Robinson (Eds.), Handbook of Logic in AI and Logic Programming, vol. 3, Clarendon Press, Oxford, 1994.
[26] J. Meheus (Ed.), Inconsistency in Science, Kluwer, Dordrecht, 2002.
[27] J. Meheus, How to reason sensibly yet naturally from inconsistencies, in: J. Meheus (Ed.), Inconsistency in Science, Kluwer, Dordrecht, 2002, pp. 151–164.
[28] F. Montagna, G. Simi, A. Sorbi, Logic and probabilistic systems, Arch. Math. Logic 35 (1996) 225–261.
[29] R.J. Parikh, Some results on the length of proofs, Trans. Amer. Math. Soc. 177 (1973) 29–36.
[30] J. Pearl, Probabilistic Reasoning in Intelligent Systems, Morgan Kaufmann, San Francisco CA, 1988.
[31] P. Pudlak, The lengths of proofs, in: S.R. Buss (Ed.), Handbook of Proof Theory, Elsevier, Amsterdam, 1998, pp. 547–637.
[32] A. Raggio, Propositional sequence calculi for inconsistent systems, Notre Dame J. Formal Logic IX (4) (1968) 359–366.
[33] D. Scott, P. Krauss, Assigning probabilities to logical formulas, in: P. Suppes, J. Hintikka (Eds.), Aspects of Inductive Logic, North-Holland, Amsterdam, 1996.
[34] C. Smorynski, The incompleteness theorems, in: J. Barwise (Ed.), Handbook of Mathematical Logic, sixth ed., North-Holland, Amsterdam, 1991, pp. 821–895.
[35] R. Solovay, Provability interpretations of modal logic, Israel J. Math. 25 (1976) 287–304.
[36] G. Takeuti, Proof Theory, North-Holland, Amsterdam, 1987.
[37] A.S. Troelstra, H. Schwichtenberg, Basic Proof Theory, Cambridge University Press, 1996.