Circumscribing Inconsistency

Philippe Besnard
IRISA
Campus de Beaulieu
F-35042 Rennes Cedex

Torsten H. Schaub*
Institut fur Informatik
Universitat Potsdam, Postfach 60 15 53
D-14415 Potsdam

Abstract

We present a new logical approach to reasoning from inconsistent information. The idea is to restore modelhood of inconsistent formulas by providing a third truth-value tolerating inconsistency. The novelty of our approach stems first from the restriction of entailment to three-valued models as similar as possible to two-valued models and second from an implication connective providing a notion of restricted monotonicity. After developing the semantics, we present a corresponding proof system that relies on a circumscription schema furnishing the syntactic counterpart of model minimization.

1 Introduction

The capability of reasoning in the presence of inconsistencies constitutes a major challenge for any intelligent system. This is because in practical settings it is common to have contradictory information. In fact, despite its many appealing features for knowledge representation and reasoning, classical logic falls in the same trap: A single contradiction may wreck an entire reasoning system, since it may allow for deriving any proposition. This comportment is due to the fact that a contradiction denies any classical two-valued model, since a proposition must be either true or false. We thus aim at providing a formal reasoning system satisfying the *principle of paraconsistency*: $\{\alpha, \neg\alpha\} \not\vdash \beta$ for some α, β. In other words, given a contradictory set of premises, this should not necessarily lead to concluding all formulas. We address this problem from a semantic point of view. We want to counterbalance the effect of contradictions by providing a third truth-value that accounts for contradictory propositions. As already put forward by [Priest, 1979], this provides us with inconsistency-tolerating three-valued models. However, this approach turns out to be rather weak in that it invalidates certain classical inferences, even if there is no contradiction. Intuitively, this is because there are too many three-valued models, in particular those assigning the inconsistency-tolerating truth-value to propositions that are unaffected by contradictions.

* Previously at LERIA, Universite d* Angers, France.

Our idea is to focus on those three-valued models that are as similar as possible to *two-valued models* of the knowledge base. In this way, we somehow hand over the model selection process to the knowledge base by preferring those models that assign *true* to as many items of the knowledge base as possible. As a result, our approach reduces nicely to classical reasoning in the absence of inconsistency. (For the reader familiar with the work of [Priest, 1989] we note that ours is different from preferring three-valued models having the highest number of classical truth-values, which amounts to approximating two-valued interpretations while somehow discarding the underlying knowledge base.) The syntactic counterpart of our preferential reasoning process is furnished by an axiom schema, similar to the ones found in circumscription [McCarthy, 1980]. Another salient feature of our approach is driven by the desire to preserve existing proofs even though they may lead to contradictory conclusions. This is because proofs provide evidence for derived conclusions. We accomplish this by introducing an implication connective that reduces (inside the knowledge base) to classical implication in the absence of inconsistency, while its resulting inferences are conserved under inconsistency.

The paper is organized as follows. Section 2 lays the semantic foundations of our approach; it presents a novel three-valued logic comprising two special connectives: The aforementioned implication and a truth-value-indicating connective (used for later axiomatization of the model selection process). To a turn, we define our *paraconsistent inference relation* by means of a preference relation over the set of models obtained in this logic. Section 3 presents the syntactic counterpart by proposing a corresponding formal proof system. We present an axiomatization of the underlying three-valued logic and we furnish a circumscription axiom providing syntactic means for reasoning from preferred inconsistency-tolerating models.

2 Model theory

This section presents our semantic approach to reasoning from possibly inconsistent knowledge bases expressed in a propositional language. We use h for classical entailment wrt two-valued interpretations and *Cn*\- for classical deductive closure. For dealing with inconsistencies we rely on an extended propositional language:

Definition 2.1 *Given a set $\mathcal{P}$ of propositional symbols let $\mathcal{L}$ be the set of all formulas generated from $\mathcal{P}$ using connectives* $\top, \bot, \neg, \vee, \wedge, \rightarrow, \leftrightarrow, \natural, \leq$.

The last two connectives serve as truth-value indicators. That is, $\natural\alpha$ means that α is true and $\alpha \leq \beta$ signifies that the truth value of α is less than that of β. This order is proper to this connective and is no intrinsic feature of the rest of the logic. We define $\top$ as $\natural\alpha \rightarrow \alpha$ and $\bot$ by $\neg\top$. Also, we define $\alpha \leftrightarrow \beta$ as $(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$. In fact, $\leq$ is also a defined connective (using $\natural$), whose discussion is deferred to Section 3.

Definition 2.2 *An interpretation is a function*

$$v : \mathcal{P} \rightarrow \{t, f, o\} \quad \textit{extending to} \quad \overline{v} : \mathcal{L} \rightarrow \{t, f, o\}$$

according to the truth tables below.

$\neg$	
t	f
f	t
o	o

$\wedge$	t	f	o
t	t	f	o
f	f	f	f
o	o	f	o

$\vee$	t	f	o
t	t	t	t
f	t	f	o
o	t	o	o

$\natural$	
t	t
f	f
o	f

$\rightarrow$	t	f	o
t	t	f	o
f	t	t	t
o	t	f	o

A model of a formula α is an interpretation that assigns either t or o to α.

Modelhood extends to sets of formulas in the standard way. Observe that $\wedge$ and $\vee$ are de Morgan duals. Also, note that the truth-value of $\alpha \rightarrow \beta$ differs from that of $\neg\alpha \vee \beta$ only in the case of $v = \{\alpha : o, \beta : f\}$ resulting in $v(\alpha \rightarrow \beta) = f$ and $v(\neg\alpha \vee \beta) = o$. This difference is prompted by the fact that t and o indicate modelhood, which motivates the assignment of the same truth-values to $\alpha \rightarrow \beta$ no matter whether we have $\alpha : t$ or $\alpha : o$. This has actually to do with the difference between *modus ponens* (MP) and *disjunctive syllogism* (DS):

$$\frac{(\alpha \rightarrow \beta) \quad \alpha}{\beta}\ \text{(MP)} \qquad\qquad \frac{(\alpha \vee \beta) \quad \neg\alpha}{\beta}\ \text{(DS)}$$

The latter yields B from $A \wedge \neg A \wedge \neg B$ because $A \vee B$ follows from A. The overall inference seems wrong because in the presence of $A \wedge \neg A$, $A \vee B$ is satisfied (by $A : o$) with no need for B to be t. This is why we center our approach upon modus ponens.

We then obtain the following consequence relation:

Definition 2.3 *Let Γ be a set of formulas and γ a formula. We define $\Gamma \Vdash \gamma$ iff each model of Γ is a model of γ.*

The reader is warned that replacement of equivalents fails: Let $\gamma[\phi_1, \ldots, \phi_k]$ be the formula obtained from $\gamma[\psi_1, \ldots, \psi_k]$ by replacing all occurrences of $\psi_1, \ldots, \psi_k$ by $\phi_1, \ldots, \phi_k$. Then,

$$\Vdash \alpha \leftrightarrow \beta \not\Rightarrow \Vdash \gamma[\alpha] \leftrightarrow \gamma[\beta]$$

Letting α be $A \rightarrow \neg\neg A$, β be $B \rightarrow \neg\neg B$, and γ be $\neg\psi$ shows the failure of replacement of equivalents: $\Vdash (A \rightarrow \neg\neg A) \leftrightarrow (B \rightarrow \neg\neg B)$ but $\not\Vdash \neg(A \rightarrow \neg\neg A) \leftrightarrow \neg(B \rightarrow \neg\neg B)$.

We now turn to the key definition of our approach:

Definition 2.4 *Let v and v' be two interpretations and Γ a set of formulas. We define*

$$v \prec_\Gamma v' \quad \textit{iff} \quad \{\gamma \in \Gamma \mid v(\gamma) = o\} \subsetneq \{\gamma \in \Gamma \mid v'(\gamma) = o\}.$$

Observe that $\prec_\Gamma$ is a strict partial order on interpretations. Hence, we can speak of minimal models for a set of formulas Γ. This leads us to the following paraconsistent inference relation:

Definition 2.5 *Let Γ be a set of formulas and γ a formula. We define $\Gamma \Vvdash \gamma$ iff each $\prec_\Gamma$-minimal model of Γ is a model of γ.*

Definition 2.4 and 2.5 show that we focus on models of Γ that assign t (instead of o) to a maximal subset of Γ. Since o accounts for inconsistency all this amounts to minimizing inconsistency. In fact, both aforementioned inference relations are paraconsistent: $\{A, \neg A\} \not\Vdash B$ and $\{A, \neg A\} \not\Vvdash B$.

Since we aim at modeling reasoning from knowledge bases expressed in a propositional language, we impose the following restriction: As modus ponens is a fairly uncontroversial reasoning mode, we take it as a basis for our approach. In particular, premises are required to be in conditional form prone to application of modus ponens:

Definition 2.6 *Let $\mathcal{P}$ be a set of propositional symbols and $\mathcal{L}_\rightarrow$ the set of all expressions of the form*

$$L_1 \wedge \ldots \wedge L_m \rightarrow L_{m+1} \vee \ldots \vee L_n$$

where $L_i \in \{\alpha, \neg\alpha \mid \alpha \in \mathcal{P}\}$ for $i = 1..n$ and $0 \leq m < n$.

We refer to expressions in $\mathcal{L}_\rightarrow$ as *clauses*. For $m = 0$, such clauses reduce to $L_1 \vee \ldots \vee L_n$. As a whole, $\mathcal{L}_\rightarrow$ is generated from $\mathcal{P}$ using connectives $\neg, \vee, \wedge, \rightarrow$.

Consider the set of formulas

$$\Gamma = \{A \rightarrow B, A \rightarrow \neg B\}. \qquad (1)$$

We obtain $\Gamma \Vvdash \neg A$. In fact, $\neg A$ is concluded for the reason that, if A were true then there would be a contradiction about B. So, when it really is the case that there *is* a contradiction about B, the reason for $\neg A$ to be concluded no longer applies. That is, we have $\Gamma \cup \{A\} \not\Vvdash \neg A$ and $\Gamma \cup \{A\} \Vvdash A \wedge B \wedge \neg B$. Observe that this example violates unrestricted monotonicity (the relative theories must be both consistent or both inconsistent; cf. Theorem 2.2). This comportment can be verified in Table 1. An entry like $o/2$ in column Γ means that interpreta-

A	B	Γ	$\Gamma \cup \{A\}$	Γ'	$\Gamma' \cup \{A\}$	Γ''	$\Gamma'' \cup \{A\}$
t	t	f	f	f	f	f	f
t	f	f	f	f	f	f	f
t	o	$o/2$	$o/2$	$o/2$	$o/2$	f	f
f	t	$t/0$	f	$t/0$	f	$t/0$	f
f	f	$t/0$	f	$t/0$	f	$t/0$	f
f	o	$t/0$	f	$t/0$	f	$t/0$	f
o	t	f	f	$o/1$	$o/2$	f	f
o	f	f	f	$o/1$	$o/2$	f	f
o	o	$o/2$	$o/3$	$o/2$	$o/3$	$o/4$	$o/5$

Table 1: Truth tables for Γ, Γ', and Γ''.

tion v, given in the first two rows, assigns o to (the conjunction of) Γ, while $|\{\gamma \in \Gamma \mid v(\gamma) = o\}| = 2$. Such a number is however just an indication and should *not* be confused with the actual ordering relation on models which is based on set inclusion! A preferred model is indicated by boldface typesetting.

For a complement, take a look at clause set

$$\Gamma' = \{\neg A \vee B, \neg A \vee \neg B\}.$$

We have $\Gamma' \cup \{A\} \not\Vdash \neg B \wedge B$. We only have $\Gamma' \cup \{A\} \Vdash B \vee \neg B$. This illustrates the difference between an implication, like $A \rightarrow B$ and a disjunction, like $\neg A \vee B$. Unlike the latter, connective $\rightarrow$ allows us to construct a proof for $B \wedge \neg B$ from $\Gamma \cup \{A\}$. Compare this with the case of all contrapositives:

$$\Gamma'' = \{A \rightarrow B, \neg B \rightarrow \neg A, A \rightarrow \neg B, B \rightarrow \neg A\}.$$

This yields $\Gamma'' \cup \{A\} \Vdash A \wedge \neg A \wedge \neg B \wedge B$.

The previous examples have illustrated that whenever there are two-valued models, they are the only *relevant* minimal models. That is, in case of a 3-valued model v assigning t, we find also all 2-valued models obtained by substituting o in v by t and f, respectively. Hence, such 3-valued models are irrelevant. See columns $\Gamma, \Gamma', \Gamma''$ in Table 1. In fact, we have the following result showing that our mechanism amounts to classical (two-valued) logic, whenever we deal with a classically consistent theory.

Theorem 2.1 *Let Γ be a classically consistent set of clauses and γ a formula whose connectives are among $\neg, \wedge, \vee$. Then,*

$$\Gamma \vdash \gamma \qquad \textit{iff} \qquad \Gamma \Vdash \gamma\,.$$

This result does not extend to the underlying inference relation $\Vdash$. A counterexample is simply Γ as in (1) and γ being $\neg A$. Also, this theorem does not extend to conclusions containing $\rightarrow$, eg. $\not\Vdash (\neg A \vee B) \rightarrow (A \rightarrow B)$ although $(\neg A \vee B) \rightarrow (A \rightarrow B)$ is a classical tautology. Theorem 2.1 is neither expected to carry over to the case where Γ is inconsistent.

A salient property of our approach is that it is *monotonic* on inconsistent premises:

Theorem 2.2 *For sets of clauses Γ and Δ, we have*

$$\Gamma \Vdash \gamma \implies \Delta, \Gamma \Vdash \gamma$$

whenever $\Gamma \Vdash \alpha \wedge \neg\alpha$ and $\forall \Gamma' \subsetneq \Gamma.\ \Gamma' \not\Vdash \alpha \wedge \neg\alpha$.

We now need a few definitions on restricted alphabets:

$$\mathcal{P}_{\natural\Gamma} = \{P \in \mathcal{P} \mid \Gamma \Vdash \natural P \vee \natural\neg P\}$$

For an alphabet $\mathcal{P}_x$, let $\mathcal{L}_x$ denote the language generated from $\mathcal{P}_x$ using $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$. Then, we have the following result showing that truthful parts of the knowledge base are closed under classical logic:

Theorem 2.3 *For all sets of clauses Γ and Δ such that $\Delta = \{\alpha \mid \Gamma \Vdash \natural\alpha\}$, we have*

$$\Delta \cap \mathcal{L}_{\natural\Gamma} = Cn_{\vdash}(\Delta) \cap \mathcal{L}_{\natural\Gamma}\,.$$

Moreover, we can show that truthful parts are never polluted by contradictions:

Theorem 2.4 *Let $\Gamma = \Gamma_1 \cup \Gamma_2$ be a clause set such that $\mathcal{P}_1 \cap \mathcal{P}_2 = \emptyset$ where $\mathcal{P}_i$ is the set of propositional symbols occurring in Γ_i. We have for each $\alpha \in \mathcal{L}_1$,*

$$\Gamma \Vdash \alpha \qquad \textit{iff} \qquad \Gamma_1 \vdash \alpha$$

whenever $\Gamma \Vdash \natural\gamma$ for each $\gamma \in \Gamma_1$.

As illustrated below, the last theorem extends in some cases to non-disjoint parts, as witnessed by $\Gamma_1, \Gamma_1^{\rightarrow}, \Gamma_1^{\leftarrow}$ below.

For further illustration, consider first the set of clauses

$$\Gamma_0 = \{\neg A, B, (\neg B \vee C)\}$$

Indeed Γ_0 has a single two-valued model $\{A : f, B : t, C : t\}$ (apart from 9 three-valued ones assigning o to Γ_0). The former is clearly the only minimal model of Γ_0. We thus have

$$\Gamma_0 \vdash \neg A \wedge B \wedge C \quad \text{and} \quad \Gamma_0 \Vdash \neg A \wedge B \wedge C.$$

Adding A to Γ_0 yields inconsistent theory $\Gamma_0' = \{A, \neg A, B, (\neg B \vee C)\}$ having only three-valued models left. In fact, all former models of Γ_0 with $A : f$ do now falsify Γ_0'. All remaining models of Γ_0 assign thus o to Γ_0' and A, the actual heart of the contradiction. Among the resulting models, we have a single minimal model, $\{A : o, B : t, C : t\}$, giving $\Gamma_0' \Vdash A \wedge \neg A \wedge B \wedge C$ by "applying" disjunctive syllogism to the consistent part of Γ_0'.

Next, consider the set of clauses

$$\Gamma_1 = \{A, \neg A, (\neg A \vee B)\}$$

This theory induces the truth-values given in Table 2. Among the three models of Γ_1, there is only one minimal one: $\{A : o, B : t\}$. As a consequence, we obtain

$$\Gamma_1 \Vdash A \wedge \neg A \wedge B\,.$$

A	B	Γ_1	$\Gamma_1^{\rightarrow}$	$\Gamma_1^{\leftarrow}$	Γ_1'	$\Gamma_1'^{\rightrightarrows}$	$\Gamma_1'^{\leftleftarrows}$	$\Gamma_1'^{\leftrightarrows}$	$\Gamma_1'^{\rightleftarrows}$
t	t	f	f	f	f	f	f	f	f
t	f	f	f	f	f	f	f	f	f
t	o	f	f	f	f	f	f	f	f
f	t	f	f	f	f	f	f	f	f
f	f	f	f	f	f	f	f	f	f
f	o	f	f	f	f	f	f	f	f
o	t	$o/2$	$o/2$	$o/2$	$o/3$	f	$o/3$	$o/3$	f
o	f	$o/3$	f	$o/3$	$o/3$	f	$o/3$	f	$o/3$
o	o	$o/3$	$o/3$	$o/3$	$o/4$	$o/4$	$o/4$	$o/4$	$o/4$

Table 2: Truth tables for Γ_1 and Γ_1'.

For those familiar with [Priest, 1989], we note that this approach has $\{A : o, B : f\}$ as a second preferred model, which denies conclusion B. See Section 4 for details. The example illustrates further the aforementioned extendibility of Theorem 2.4: Despite the inconsistency of A, we derive B from the consistent premises A and $\neg A \vee B$.

Actually, things do not necessarily change by orienting the above disjunctions as implications:

$$\Gamma_1^{\rightarrow} = \{A, \neg A, (A \rightarrow B)\} \text{ and } \Gamma_1^{\leftarrow} = \{A, \neg A, (\neg B \rightarrow \neg A)\}$$

$\Gamma_1^{\rightarrow}$ and $\Gamma_1^{\leftarrow}$ have the same minimal model as Γ_1; thus offering the same conclusions. However, while $\Gamma_1^{\leftarrow}$ has the same models as Γ_1, interpretation $\{A : o, B : f\}$ falsifies $\Gamma_1^{\rightarrow}$.

Adding clause $A \vee \neg B$ to Theory Γ_1 yields

$$\Gamma_1' = \{A, \neg A, (\neg A \vee B), (A \vee \neg B)\}$$

Γ_1' has two minimal models, both of which were models of Γ_1, yet only one of them was Γ_1-preferred. We thus get

$$\Gamma_1' \Vvdash A \wedge \neg A \quad \text{and} \quad \Gamma_1' \not\Vvdash B$$

illustrating that inferences by disjunctive syllogism are not always preserved.

For a complement, consider rule sets

$$\begin{aligned}
\Gamma_1'^{\rightrightarrows} &= \{A, \neg A, A \rightarrow B, \neg A \rightarrow \neg B\} \\
\Gamma_1'^{\leftleftarrows} &= \{A, \neg A, \neg B \rightarrow \neg A, B \rightarrow A\} \\
\Gamma_1'^{\rightleftarrows} &= \{A, \neg A, A \rightarrow B, B \rightarrow A\} \\
\Gamma_1'^{\leftrightarrows} &= \{A, \neg A, \neg B \rightarrow \neg A, \neg A \rightarrow \neg B\}
\end{aligned}$$

From these, we obtain after consulting Table 2:

$$\begin{aligned}
\Gamma_1'^{\rightrightarrows} &\Vvdash A \wedge \neg A \wedge B \wedge \neg B \\
\Gamma_1'^{\leftleftarrows} &\Vvdash A \wedge \neg A \\
\Gamma_1'^{\rightleftarrows} &\Vvdash A \wedge \neg A \wedge B \\
\Gamma_1'^{\leftrightarrows} &\Vvdash A \wedge \neg A \wedge \neg B
\end{aligned}$$

The derivability of B and $\neg B$ illustrates the role of connective $\rightarrow$ as proof-provider: All proofs obtained from clauses by modus ponens are set in stone. This general property is reflected by the validity of $\Vdash (\alpha \wedge (\alpha \rightarrow \beta)) \rightarrow \beta$.

3 Proof theory

This section presents a formal proof system for our approach to circumscribing inconsistency. In analogy to the semantics, we first axiomatize II- and then we account for minimization by providing a syntactic axiom schema, so that the resulting system axiomatizes II-

The axiomatization of Ih consists of modus ponens as inference rule and the following axiom schemas:

$$\alpha \vee \neg\alpha \quad (1)$$
$$\alpha \wedge \beta \rightarrow \alpha,\ \alpha \wedge \beta \rightarrow \beta \quad (2)$$
$$\alpha \rightarrow \alpha \vee \beta,\ \alpha \rightarrow \beta \vee \alpha \quad (3)$$
$$\alpha \rightarrow (\beta \rightarrow (\alpha \wedge \beta)) \quad (4)$$
$$((\alpha \rightarrow \beta) \rightarrow \alpha) \rightarrow \alpha \quad (5)$$
$$(\alpha \rightarrow \gamma) \rightarrow ((\beta \rightarrow \gamma) \rightarrow (\alpha \vee \beta \rightarrow \gamma)) \quad (6)$$

$$\alpha \leftrightarrow \neg\neg\alpha \quad (7)$$
$$\neg(\alpha \vee \beta) \leftrightarrow \neg\alpha \wedge \neg\beta \quad (8)$$
$$\neg(\alpha \wedge \beta) \leftrightarrow \neg\alpha \vee \neg\beta \quad (9)$$

$$\alpha \rightarrow (\beta \rightarrow \alpha) \quad (10)$$
$$(\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma)) \quad (11)$$
$$(\alpha \rightarrow \beta) \rightarrow \neg\alpha \vee \beta \quad (12)$$
$$\alpha \wedge \neg\beta \rightarrow \neg(\alpha \rightarrow \beta) \quad (13)$$

$$\natural\alpha \rightarrow \alpha \quad (14)$$
$$\natural\alpha \rightarrow \natural\natural\alpha \quad (15)$$
$$\natural\neg\alpha \rightarrow \neg\natural\alpha \quad (16)$$
$$\neg\natural\alpha \leftrightarrow \natural\neg\natural\alpha \quad (17)$$
$$\natural(\alpha \wedge \beta) \leftrightarrow \natural\alpha \wedge \natural\beta \quad (18)$$
$$\natural(\alpha \vee \beta) \leftrightarrow \natural\alpha \vee \natural\beta \quad (19)$$
$$\natural(\alpha \rightarrow \beta) \rightarrow (\natural\alpha \rightarrow \natural\beta) \quad (20)$$

$$\neg\alpha \rightarrow \natural(\alpha \rightarrow \beta) \quad (21)$$
$$\natural(\neg(\alpha \rightarrow \beta)) \rightarrow \alpha \wedge \natural\neg\beta \quad (22)$$
$$\natural((\alpha \rightarrow \beta) \rightarrow \alpha) \rightarrow \alpha \quad (23)$$

$$\natural\alpha \leftrightarrow \natural\beta \quad \text{for} \quad \alpha \leftrightarrow \beta \in \{(7), (8), (9)\} \quad (24)$$
$$\natural\alpha \rightarrow \natural\beta \quad \text{for} \quad \alpha \rightarrow \beta \in \{(10), \ldots, (13)\} \quad (25)$$
$$\natural\alpha \quad \text{for} \quad \alpha \in \{(14), \ldots, (20)\} \quad (26)$$

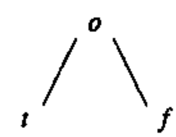
As can be shown, this proof system is sound and complete for $\Vdash$. We write $\gamma \in Cn_{\Vdash}(\Gamma)$ to indicate that γ can be derived from Γ by the above proof system.

Semantically, the move from I- to II- amounts to minimizing the set of premises with truth-value *o*. That is, we prefer models that assign truth-value *o* to a minimal set of premises. We can turn this idea into the syntax by using a connective indicating that a formula has a truth value which is less than the one of another formula. As anticipated in Section 2, such a connective can be defined as follows:

$$\alpha \leq \beta =_{\text{def}} (\natural\alpha \wedge \natural\beta) \vee (\natural\neg\alpha \wedge \natural\neg\beta) \vee (\neg\natural\beta \wedge \neg\natural\neg\beta)$$

This induces the following truth table corresponding to the poset of truth-values on the right hand side.

$\leq$	t	f	o
t	t	f	t
f	f	t	t
o	f	f	t



With this connective, we are now ready to express the following *circumscription schema* providing a syntactic account for preferring $\prec_\Gamma$-minimal models. For readability, we identify in the next definition clause set $\{\gamma_1, \ldots, \gamma_n\}$ with $\bigwedge_{i=1}^n \gamma_i$.

Definition 3.1 *Let $\Gamma = \{\gamma_1, \ldots, \gamma_n\}$ be a finite set of formulas over alphabet $P_1, \ldots, P_k$ (abbreviated $\vec{P}$) so that $\gamma_i = \gamma_i[\vec{P}]$ and $\Gamma = \Gamma[\vec{P}]$. We define the three-valued paraconsistent circumscription schema $Circ_3 P(\Gamma)$ as*

$$\Gamma[\vec{\phi}] \wedge \left(\bigwedge_{i=1}^n \gamma_i[\vec{\phi}] \leq \gamma_i[\vec{P}]\right) \rightarrow \left(\bigwedge_{i=1}^n \gamma_i[\vec{P}] \leq \gamma_i[\vec{\phi}]\right)$$

Importantly, combining $Circ_3 P(\Gamma)$ with the proof system for $\Vdash$ captures the desired paraconsistent inference relation $\Vvdash$:

Theorem 3.1 *Let Γ be a set of clauses. Then, we have*

$$\Gamma \Vvdash \gamma \quad \textit{iff} \quad \gamma \in Cn_{\Vdash}(\Gamma \cup \{Circ_3 P(\Gamma)\})$$

For illustration, let us return to our initial example

$$\Gamma = \{A \rightarrow B, A \rightarrow \neg B\}$$

We consider the following instance of $Circ_3 P(\Gamma)$ where $\phi_A = \bot$ and $\phi_B = B$:

$$\begin{array}{ll} & ((\bot \rightarrow B) \wedge (\bot \rightarrow \neg B)) \\ \wedge & ((\bot \rightarrow B \leq A \rightarrow B) \wedge (\bot \rightarrow \neg B \leq A \rightarrow \neg B)) \\ \rightarrow & ((A \rightarrow B \leq \bot \rightarrow B) \wedge (A \rightarrow \neg B \leq \bot \rightarrow \neg B)) \end{array}$$

From Γ, we obtain the right hand side (RHS) of $Circ_3 P(\Gamma)$, that is, $(A \rightarrow B \leq \bot \rightarrow B)$ and $(A \rightarrow \neg B \leq \bot \rightarrow \neg B)$ after establishing the LHS by means of theorem $(\alpha \wedge (\beta \leq \top)) \rightarrow (\beta \leq \alpha)$. By applying transitivity of $\leq$ to RHS and $(\bot \rightarrow B) \leq \top$ and $(\bot \rightarrow \neg B) \leq \top$, we then get $(A \rightarrow B) \leq \top$ and $(A \rightarrow \neg B) \leq \top$. So, we get $(\neg A \leq \top) \vee ((\neg B \wedge B) \leq \top)$ yielding $\neg A \leq \top$, hence $\neg A$. Notably, it is the circumscription schema that reduces the three-valued consequence relation $\Vdash$ to its classical two-valued counterpart $\vdash$ (cf. Theorem 2.1).

For further illustration, consider $\Gamma \cup \{A\}$ along with the instance of $Circ_3 P(\Gamma \cup \{A\})$ obtained by taking $\phi_A = \top$ and $\phi_B = B \wedge \neg B$. We obtain $A \leq \top$ and so A using theorem $(\alpha \wedge (\gamma \leq \beta)) \rightarrow (\gamma \leq (\alpha \rightarrow \beta))$. Of course, not every $\Vdash$ conclusion necessitates the circumscription schema in order to be derived. For instance, B and $\neg B$ are directly derived by modus ponens from $\Gamma \cup \{A\}$.

4 Related work

There are a number of proposals addressing inconsistent information. At first, there is the wide range of paraconsistent logics [Priest *et al.*, 1989]. As opposed to our approach, such logics usually fail to identify with classical logic when the set of premises is consistent. There are also many approaches dealing with classical reasoning from consistent subsets. In a broader sense, this includes also belief revision and truth maintenance systems. A comparative study of the aforementioned approaches in general is given in [Besnard, 1991].

A system, at first sight closely related to ours, is LP_m [Priest, 1989]; it was conceived to overcome the failure of disjunctive syllogism in LP [Priest, 1979]. LP amounts to the 3-valued logic obtained by restricting $\Vdash$ to connectives $\neg, \vee$ and $\wedge$ and defining $\alpha \rightarrow \beta$ as $\neg\alpha \vee \beta$. In LP_m modelhood is then limited to models containing a minimal number of *prepositional variables* being assigned *o*. As our approach, this allows for drawing "all classical inferences except where inconsistency makes them doubtful anyway" [Priest, 1989]. There are two major differences though: First, the aforementioned restriction of modelhood focuses on models as close as possible to 2-valued interpretations, while the one in our approach aims at models next to 2-valued *models* of the considered formula. The effects of making the formula select its preferred models can be seen by looking at Γ_1: While LP_m yields two preferred models $\{A : o, B : t\}$ and $\{A : o, B : f\}$ from which one obtains $A \wedge \neg A, \Gamma_1$ makes our approach prefer the former over the latter, thus yielding B as additional conclusion. Second, we have introduced implication as a primitive connective rather than a defined one. As a consequence, a modus ponens inference, like deriving B from A and $A \rightarrow B$, is preserved no matter what other premises are given; this fails in LP_m. Note that we get distinct truth-tables (and so different conclusions) for Γ_1' and its variants $\Gamma_1'^{\rightleftarrows}, \ldots$, while LP_m does not differentiate these variations. A resolution-based system close to LP yet with a stronger disjunction is described in [Lin, 1987].

A whole variety of approaches uses lattices for dealing with inconsistency, eg. [Arieli and Avron, 1994; Belnap, 1977; Sandewall, 1985]. For instance, [Arieli and Avron, 1994; 1996] describes a system based on 4-valued logic that allows for constraining "the most consistent" models in the meta-level by a user-given set of propositions taking classical truth-values only. [Carnielli *et al.*, 1991] proposes a translation-based approach to reasoning in the presence of contradictions that translates a logic into a family of other logics, eg. classical logic into 3-valued logics.

The difference between our approach and "reasoning from maximal consistent subsets of the premises" is that we still pay attention to one objection motivating relevant logics [Anderson and Belnap, 1975] and that is applying disjunctive syllogism to contradictory premises. However, we do not go as far as sanctioning any classical inference not using inconsistent subformulas. That is, we still follow the principle of relevant logics that an inference rule is a priori applicable to any premise. This is in contrast with the idea of restricted access logic [Gabbay and Hunter, 1993], where all classical inference rules are admitted with some special application conditions.

Among others, logic programming with inconsistencies was addressed in [Blair and Subrahmanian, 1988; 1989]. [Wagner, 1991] describes a procedural framework for handling contradictions that relies on the notions of "support" and "acceptance". The former avenue of research is further developed in [Grant and Subrahmanian, 1995], where it is shown how the approach of [Blair and Subrahmanian, 1988] can be extended by classical inferences, like reasoning by cases. Intuitively, the corresponding entailment relations amount to logic programming in a 3-valued (and 4-valued, respectively) logic. The major difference to our approach is that compared to classical entailment, these approaches are sound but not complete (even when the set of premises is consistent). As with other approaches, this is because they aim at paraconsistent reasoning in a logic programming setting that does not necessarily coincide with classical logic.

Our approach is clearly semantical in contrast to many other proposals to paraconsistency: (i) the idea of "forgetting" literals [Kifer and Lozinskii, 1989; Besnard and Schaub, 1996]; (ii) the idea of stratified theories [Benferhat *et al.*, 1993]; (iii) the idea of reliability relation [Roos, 1992], (iv) and more generally the idea of reasoning from consistent subsets of the premises. In contrast to [Tlirner, 1990], where the baseline is to analyze propositions (so as to resolve paradoxes about truth, for instance), we simply apply a system of truth-values so that we can have non-trivial inconsistent premises. Moreover, our approach is purely deductive, as opposed to argumentation-based frameworks, like [Wagner, 1991; Elvang and Hunter, 1995]. An unusual approach to reasoning from inconsistency is due to [Lin, 1996], who introduces the notion of consistent belief by means of modal operators. This approach fails to satisfy reflexivity (not every premise is concluded).

5 Conclusion

We presented a semantical approach to dealing with inconsistent knowledge bases that is founded on the minimization of three-valued models. This was complemented by a formal proof system accomplishing model minimization by appeal to a circumscription axiom. The distinguishing features of our approach are (i) its desire to provide models making true (instead of true and false) as many as possible items of the knowledge base, (ii) its centering on inferences drawn by modus ponens by means of a primitive implication connective, and (iii) its property of restricted monotonicity. A major further development will be lifting the approach to the first-order case. In this context, we draw the reader's attention to the fact that our approach (unlike [Priest, 1989]) does not rely on the notion of an atomic proposition, which is always problematic when passing from the propositional case to the first-order case.

References

[Anderson and Belnap, 1975] A. Anderson and N. Belnap. *Entailment: The Logic of Relevance and Necessity.* Princeton University Press, 1975.

[Arieli and Avron, 1994] O. Arieli and A. Avron. Logical bilattices and inconsistent data. In *Logic in Computer Science Conf,* pp 468-476,1994.

[Arieli and Avron, 1996] O. Arieli and A. Avron. Automatic diagnoses for properly stratified knowledge-bases. In *Int. Conf. on Tools with Artificial Intelligence,* pp 392-399. IEEE Press, 1996.

[Belnap, 1977] N. Belnap. A useful four-valued logic. In J. Dunn and G. Epstein, eds, *Modern Uses of Multiple-Valued Logic.* Reidel, 1977.

[Benferhat et al., 1993] S. Benferhat, D. Dubois, & H. Prade. Argumentative inference in uncertain and inconsistent knowledge bases. In *Int. Conf on Uncertainty in Artificial Intelligence,* pp 411-419,1993.

[Besnard and Schaub, 1996] P. Besnard and T. Schaub. A simple signed system for paraconsistent reasoning. In *European Workshop on Logics in Artificial Intelligence,* pp 404-416. Springer Verlag, 1996.

[Besnard, 1991] P. Besnard. Paraconsistent logic approach to knowledge representation. In *World Conf on Fundamentals of Artificial Intelligence,* 1991.

[Blair and Subrahmanian, 1988] H. Blair and V.S. Subrahmanian. Paraconsistent foundations of logic programming." *Journal of Non-Classical Logics,* 5(2):45-73,1988.

[Blair and Subrahmanian, 1989] H. Blair and V.S. Subrahmanian. Paraconsistent logic programming. *Theoretical Computer Science,* 68(2):135-154,1989.

[Carnielli *etal,* 1991] W. Carnielli, L. Fariftas del Cerro, and M. Lima Marques. Contextual negations and reasoning with contradictions. In *Int. Joint Conf on Artificial Intelligence,* pp 532-537. Morgan Kaufmann, 1991.

[Elvang and Hunter, 1995] M. Elvang and A. Hunter. Argumentative logics: reasoning with classically inconsistent information. *Journal of Knowledge and Data Engineering,* 16:125-145,1995.

[Gabbay and Hunter, 1993] D. Gabbay and A. Hunter. Restricted access logics for inconsistent information. In *European Conf on Symbolic and Quantitative Approaches to Reasoning and Uncertainty.* Springer Verlag, 1993.

[Grant and Subrahmanian, 1995] J. Grant and V.S. Subrahmanian. Reasoning in inconsistent knowledge bases. *IEEE Transactions on Knowledge and Data Engineering,* 7(1):177-189,1995.

[Kifer and Lozinskii, 1989] M, Kifer and E. Lozinskii. RI: A logic for reasoning with inconsistency. In *Logic in Computer Science,* pp 253-262,1989.

[Lin, 1987] F. Lin. Reasoning in the presence of inconsistency. In *AAAI Nat. Conf on Artificial Intelligence,* pp 139-143. AAAI/MIT Press, 1987.

[Lin, 1996] J. Lin. A semantics for reasoning consistently in the presence of inconsistency. *Artificial Intelligence,* 86(1-2):75-95, 1996.

[McCarthy, 1980] J. McCarthy. Circumscription — a form of nonmonotonic reasoning. *Artificial Intelligence,* 13(1-2):27-39,1980.

[Priest *etal,* 1989] G. Priest, R. Routley, and J. Norman, editors. *Paraconsistent Logics.* Philosophica Verlag, 1989.

[Priest, 1979] G. Priest. Logic of paradox. *Journal of Philosophical Logic,* 8:219-241,1979.

[Priest, 1989] G. Priest. Reasoning about truth. *Artificial Intelligence,* 39:231-244,1989.

[Roos, 1992] N.Roos. A logic for reasoning with inconsistent knowledge. *Artificial Intelligence,* 57:69-103,1992.

[Sandewall, 1985] E. Sandewall. A functional approach to non-monotonic logic. *Computational Intelligence,* 1:80-87,1985.

[Turner, 1990] R. Tbrner. *Truth and Modality for Knowledge Representation.* Pitman, 1990.

[Wagner, 1991] G. Wagner. Ex contradictione nihil sequitur. In *Int. Joint Conf on Artificial Intelligence,* pp 538-543. Morgan Kaufmann, 1991.