UNIVERSITÄT DORTMUND
FACHBEREICH INFORMATIK

LEHRSTUHL VIII
KÜNSTLICHE INTELLIGENZ

A Note on Paraconsistent Entailment in Machine Learning

Preliminary Version

LS–8 Report 10

Steffo Weber and Siegfried Bell
bell,steffo@ls8.informatik.uni-dortmund.de

Dortmund, April 13, 1994

Forschungsberichte des Lehrstuhls VIII (KI),
Fachbereich Informatik
der Universität Dortmund

ISSN 0943-4135

Anforderungen an:

Universität Dortmund
Fachbereich Informatik
Lehrstuhl VIII
D-44221 Dortmund

Research Reports of the unit no. VIII (AI)
Computer Science Department
of the University of Dortmund

ISSN 0943-4135

Requests to:

University of Dortmund
Fachbereich Informatik
Lehrstuhl VIII
D-44221 Dortmund

e-mail: reports@ls8.informatik.uni-dortmund.de
ftp: kimo.informatik.uni-dortmund.de:pub/Reports

A Note on Paraconsistent Entailment in Machine Learning

Preliminary Version

LS–8 Report 10

Steffo Weber and Siegfried Bell
`bell,steffo@ls8.informatik.uni-dortmund.de`

Dortmund, April 13, 1994

Universität Dortmund
Fachbereich Informatik

Abstract

Recent publications witness that there is a growing interest in multi-valued logics for machine learning; some of them arose as a more or less formal description of a computer program's inferential behaviour. The referred origin of these systems is Belnap's four-valued logic, which has been adopted for the various needs of knowledge representation in a machine learning system. However, it is unclear what an inconsistent knowledge base entails. We investigate MOBAL's logic $\Re$ and show how to interpret the term 'paraconsistent inference' of this system. It turns out that the meaning of the basic connective $\rightarrow$ of $\Re$ can be represented as a combination of two systems of Kleene's strong three-valued logic, where the two systems differ in the set of designated truth values. The resulting logic is functionally complete but the entailment relation is not axiomatizable. This drawback yields a fundamental difference between nonmontonicity within belief-revision and non-monotonic reasoning systems like Servi's refinement γ_1 of Gabbay's γ.

1 Paraconsistency

Paraconsistency is a general term to describe the property of a logic where the principle of *ex falso sequitur quodlibet* (EFQ) fails, that is, something inconsistent[1] does *not* entail the set of all well formed formulas (wffs). To see how EFQ comes into play in standard logics, consider the following example: suppose, we know $\{A, \neg A\}$. Then, by OR-introduction this set entails $\{A, \neg A \vee B\}$ for an arbitrary sentence B and by the principle of disjunctive syllogism, we can conclude B from $\{A, \neg A \vee B\}$.

How can we avoid this fallacy? In general there are two possibilities. The first one tries to 'clear' the inconsistent set of formulas, and the keyword is *belief revision*. The second one allows inconsistencies in the sense of a paraconsistent logic. In order to formulate a paraconsistent consequence relation we have to find a model for $\{A, \neg A\}$. Well, obviously it is no problem to define an appropriate truth table where $\{A, \neg A\}$ is satisfiable, but what does this really *mean*? Belnap ([Belnap, 1977]) argues that we should not interpret the sentence A ontologically but epistemically. That is, if an interpretation function assigns the truth value *true* to A, then this does not mean 'A is true' but 'we have been *told* that A is true'. Hence, $\{A, \neg A\}$ means that we have been told (maybe by two different persons) both, that A holds and that $\neg A$ holds.

Our approach is to come up with a four-valued semantics, i.e., the set of truth values is $\tau = \{t, f, \top, \bot\}$. If I is an interpretation function, A an atomic sentence we write $I(A) = \top$ in order to denote 'A and $\neg A$ hold' and $I(A) = \bot$ to denote 'we do not have any information about A', that is, it is unknown whether A holds or not. If a sentence has the epistemic truth value $\top$ or $\bot$, then its ontological status is uncertain which does not mean that it will stay uncertain in eternity. A sentence whose truth value is $\top$ is overdetermined, in the sense that we have too much data (and less information) concerning A and if its truth value is $\bot$ then it is underdetermined. Note that the different interpretations of the third truth value in some three-valued logics correspond to $\top$ and $\bot$[2].

1.1 MOBAL from a logical point of view

MOBAL's inference engine [Morik et al.,1933] and its underlying logic $\Re$ try to accomodate several aspects of paraconsistent reasoning, which will be shown to be incompatible. The very postulate of $\Re$ is that classical inconsistencies are allowed, in the sense that they should be entailed (i.e. $\{A, B, A \rightarrow \neg B\} \vdash \neg B$), but from something inconsistent nothing should be inferred (i.e. $\{A, \neg A, A \rightarrow B\} \not\vdash B$). These two principles are incompatible. In order to see this let us state them more formally. Let Cn be a logical consequence operator and $\vdash$ a corresponding provability relation. The following is a wishlist for Cn.

1. $Cn(\{A, \neg A\}) \neq$ *Form*(Σ), where *Form*(Σ) is the set of all well-formed formulas w.r.t. to a signature Σ.

[1] Throughout the paper we will use the term 'inconsistency' in order to describe contrary information, which is the same in classical two-valued logic but in a multiple-valued logic we could have contrary information within a set of sentences although there is a model for these sentences.

[2] According to Rescher ([Rescher, 1969]), Bochvar interprets his third truth value as 'paradoxical', 'meaningless' or 'undecidable' to overcome difficulties that arise with antinomies whereas Kleene ([Kleene, 1952]) interprets the third value as a truth value gap. Another possibility is a partial interpretation where some sentences simply don't have a truth value.

2. From something 'inconsistent' nothing should be inferred:

$$\frac{\vdash A, A \rightarrow B \qquad \nvdash \neg A}{\vdash B}$$

3. 'Inconsistencies' should be allowed:

$$\frac{\vdash A, \neg B, A \rightarrow B}{\vdash B}$$

The formalization of item 2) actually reads 'conclude only from sentences which do not contain contrary information' which is a rewording of the item's informal description. Item 3) is a punctualization of 1). If one is willing to accept two other rules (contraposition and $\rightarrow$-introduction) it is easy to show that the items on our wishlist are incompatible. Since it is claimed by [Morik et al.,1933], p. 41, that the connectives $\wedge$ and $\rightarrow$ have the usual meaning (and therefore the usual properties) there is no reason to reject the following rules of inference:

$$\frac{\vdash A \rightarrow B}{\vdash \neg B \rightarrow \neg A}\ (CP) \qquad \frac{\begin{array}{c}[\vdash A] \\ \vdots \\ \vdash B\end{array}}{\vdash A \rightarrow B}\ (\rightarrow I)$$

Now let us examine item 2 of the wishlist, which is our formulation of 'from something inconsistent, nothing should be inferred'. What is meant by this statement is that if both A and $\neg A$ are known as well as $A \rightarrow B$, then B should not be entailed. This non-monotonic effect was originally called 'blocking' by Wrobel (in [Morik et al.,1933]). A special case may occur if we consider the set $\{A, A \rightarrow \neg A\}$. The problem is that within such a circularity and according to item 2 the derivation of $\neg A$ could only be blocked, if it was permitted in a previous step. Hence, $\{A, A \rightarrow \neg A\}$ should not entail $\neg A$ although this cannot be guaranteed by the rule in item 2. On way to hack around this problem is to reformulate item 2 as

$$\frac{\vdash A, A \rightarrow B \qquad \nvdash \neg A \text{ and } B \not\equiv \neg A}{\vdash B}$$

where $\equiv$ stands for an appropriate notion of semantical equivalence.

The insufficiency of this formulation can be seen by a counter-example because the above pattern of odd loops can be applied in a more general form: $X = \{A, A \rightarrow B, B \rightarrow \neg A\} \nvdash \neg A$. A solution could be found if we had the following rule of inference

$$\frac{\vdash A \rightarrow B, B \rightarrow C}{\vdash A \rightarrow C}$$

Together with $(\rightarrow I)$ we can derive $A \rightarrow \neg A$ from $\{A, \neg A\}$ and $A \rightarrow \neg A$ from $\{A \rightarrow B, B \rightarrow \neg A\}$. To capture this strange kind of loops properly within a non-monotonic inference rule, we introduce *non-monotonic MP (NMP)*[3], which does not only check whether $\neg A$ is not derivable but if $\nvdash A \rightarrow \neg A$.

[3] This notion mirrors exactly a difficulty by which all nonmonotonic formalisms are plagued (as stated in [Servi, 1992]).

$$\frac{\vdash A, A \rightarrow B \qquad \nvdash A \rightarrow \neg A}{\vdash B} \textit{(NMP)}$$

We think that this is the correct formalization of 'from something inconsistent nothing should be inferred'. We now show the incompatibility between NMP and item 3) of our wishlist.

$$\begin{array}{rcl} X = \{A, \neg B, A \rightarrow B\} & \vdash & A \rightarrow \neg B \text{ by } (\rightarrow I) \\ & \vdash & B \rightarrow \neg A \text{ by contraposition} \\ & \vdash & A \rightarrow \neg A \text{ by transitivity} \end{array}$$

That is, there is a conflict between 'from something inconsistent, nothing should be inferred' and 'inconsistencies are allowed'.

2 Semantical investigation

In the last section we have seen that the main aspects of MOBAL's concept of paraconsistency – using a straightforward formulation – are incompatible. In order to make a new wishlist, we have to point out which logical preliminaries we accept. These preliminaries concern basic properties of the entailment relation as well as the logical connectives. Our requirements concerning these two points will be put into a formal framework in the next section, therefore we want to leave some things intuitive for now. As a language we assume atomic propositions, negation ($\neg$), disjunction ($\vee$) and implication ($\rightarrow$). A formula can take the truth value overdetermined ($\top$), undefined ($\perp$) true (t) or false (f). As designated truth values we take t and $\top$. Note that these imply a concept of entailment based on the standard notion of Bolzano.

For the entailment relation we have the following properties in mind: first, it should be reflexive, i.e.

$$\frac{\vdash A}{\vdash A} \text{ Reflexivity}$$

Cumulativity is needed to guarantee the order-independence of entailment steps, that is if X entails A then the addition of A to X does not block the entailment of an arbitrary sentence B entailed by X (hence, causes no additional non-monotonic effect).

$$\frac{X \vdash A \quad X \vdash B}{X \cup A \vdash B} \text{ Cumulativity}$$

For the connectives $\neg, \vee, \rightarrow$: we wish them to be as close as possible to classical logic. That is, we want to have double negation, material implication and the usual definition of satisfiability for $\vee$ (a disjunct statement is satisfied if one of its disjuncts is satisfied). Additionally we have to give up one of the most important properties of classical logic, the principle of tertium non datur. To pretty print these thoughts

$$\frac{\vdash \neg\neg A}{\vdash A} DN \qquad \frac{\vdash A \rightarrow B}{\vdash \neg A \vee B} MI \qquad \frac{\vdash A \vee B}{\vdash B \vee A} \qquad \frac{\vdash A \vee A}{\vdash A}$$

As a derived rule, we have contraposition. The invalidity of tertium non datur can only be stated semantically. Now let's focal the real point of paraconsistent reasoning. If we take $\rightarrow$ as material implication, the sentence $A \rightarrow \neg A$ reduces to $\neg A \vee \neg A$ which equals $\neg A$. Since our entailment relation should be reflexive, we have

$$\frac{\vdash A, A \rightarrow \neg A}{\vdash \neg A}$$

At a first glance this looks strange, since there is an inference from something 'inconsistent' via modus ponens but a closer look shows, that $\neg A$ was given explicitely by the assertion of $A \rightarrow \neg A$.

Another point on our former wishlist was the avoidance of odd loops, i.e.,

$$\frac{\vdash A \rightarrow B, \quad B \rightarrow \neg A}{\not\vdash A \rightarrow \neg A}$$

Now, if we transform the two implications in the premise, we get $\{\neg A \vee B, \neg B \vee \neg A\}$ from which we cannot conclude $\neg A$, since B could be overdetermined and therefore $\neg A$ is not entailed. This sounds reasonable and is very close to the behaviour of MOBAL. But the crucial point is: can we expect $X = \{A, \neg B, A \rightarrow B\}$ to entail B? Clearly, X entails $\neg B \rightarrow \neg A$ by contraposition. If we allow the derivation of B, then we must allow the derivation of $\neg A$. But then we would have applied modus ponens to $A, A \rightarrow B$ although A and $\neg A$ holds. Hence, neither $\neg A$ nor B should be entailed by X. To see that this argument is consistent with our semantical view, transform $A \rightarrow B$ to $\neg A \vee B$ which yields $X' = \{A, \neg B, \neg A \vee B\}$. A could be overdetermined as well as B. If A is overdetermined, $\neg A \vee B$ is satisfied but not B, and if B is overdetermined the disjunction is satisfied but not $\neg A$. Hence, neither B nor $\neg A$ should be entailed by X'.

The last point shows that if we accept a few properties of our entailment relation (Inclusion, Cumulativity) and a few standard characteristics of the logical connectives $\neg, \vee$ and $\rightarrow$, the aims of MOBAL's paraconsistent logic

- inconsistencies should be allowed and inconsistent conclusions should be drawn (via a restricted version of MP) and
- from something inconsistent no conclusion should be drawn via MP.

are still incompatible.

2.1 Truth tables

Up to now we have mentioned paraconsistency and the behaviour of $\top$ in an inference. But nothing has been said about the fourth truth value, $\bot$, except that it means something like 'underdetermined' and does not belong to the set of designated truth values. In this section we develop a complete truth table for the basic connectives $\wedge$ (conjunction) and $\neg$ (negation), from which $\vee$ and $\rightarrow$ can be defined[4]. From these tables we obtain an entailment relation, which meets our aims, defined in the previous section.

[4] In the rest of the paper we will treat $\rightarrow$ and $\neg$ as basic connectives. Notwithstanding, for didactical and illustrative reasons we will use in the current section $\neg$ and $\wedge$ as basic.

Consider the following set of sentences $X = A, \neg A, A \wedge B$. The interpretation that A is both *true and false* is necessary in order to find a model for X. But then, what is the truth value of a compound statement like $A \wedge B$? That is we have to fill the gaps in the following truth value table:

$\wedge$	t	f	$\top$	$\perp$
$\top$	?	?	?	?

Our interpretation is that if a sentence A has the truth value $\top$ or $\perp$ then it is (up to now) uncertain, but we will see in the future whether the sentence is true or false. Therefore, suppose what will happen to $A \wedge B$ if $I(A) = \top$ and $I(B) = f$. In this case A could become true or false but due to $I(B) = f$ the sentence $A \wedge B$ will never become true. Hence,

$\wedge$	t	f	$\top$	$\perp$
$\top$	?	f	?	?

The case $I(B) = t$ is similiar. Here the truth value of the conjunct depends solely on the truth value of A which is the uncertain truth value $\top$. Therefore,

$\wedge$	t	f	$\top$	$\perp$
$\top$	$\top$	f	?	?

What about the remaining gaps? Their value must be an uncertain one like $\top$ or $\perp$. We think that a sentence whose truth value is $\top$ should be read as *overdetermined* (*underdetermined* in the case that $I(A) = \perp$). And if A and B are overdetermined so is the adjunct $A \wedge B$.

$\wedge$	t	f	$\top$	$\perp$
$\top$	$\top$	f	$\top$	?

The last gap is the most difficult one and we've got no knocking down philosophical arguments why it should be $\perp$ except the following one: if, and the thing wildly possible, we want $\{A \wedge B\}$ to entail A and to entail B, we have to prevent that $I(A) = \top$ and $I(B) = \perp$ satisfies $\{A \wedge B\}$ (keeping in mind that $\top$ is a designated truth value).[5] This leads us to

$\wedge$	t	f	$\top$	$\perp$
$\top$	$\top$	f	$\top$	$\perp$

Similiar arguments are leading to the cases where $I(A) = \perp$. If B is true, then the truth of $A \wedge B$ depends merely on A.

[5] Note that there is an argument why the truth value of the adjunct should be $\top$: if something has neither been told true nor false (that is it is unknown) the assertion of such a sentence is consistent in a classical sense. That is any sentence whose truth value is $\perp$ can be asserted without causing (big) problems. But if we know that of a conjunctive sentence $A \wedge B$ one part is $\top$ while the other one is $\perp$, then we cannot consistently assume that $A \wedge B$ holds.

$\wedge$	t	f	$\top$	$\perp$
$\top$	$\top$	f	$\top$	$\perp$
$\perp$	$\perp$	?	?	?

Now if B is false, then the compound statement will never become true. That is,

$\wedge$	t	f	$\top$	$\perp$
$\top$	$\top$	f	$\top$	$\perp$
$\perp$	$\perp$	f	?	?

If B is $\top$ and we want $\wedge$ to be an associative operation, we can reduce this to the case where $I(A) = \top$ and $I(B) = \perp$.

$\wedge$	t	f	$\top$	$\perp$
$\top$	$\top$	f	$\top$	$\perp$
$\perp$	$\perp$	f	$\top$	?

The case where $I(A) = I(B) = \perp$ should be read as A is underdetermined and B is underdeterminded. Hence, $A \wedge B$ is underdetermined.

The complete table is:

$\wedge$	t	f	$\top$	$\perp$
t	t	f	$\top$	$\perp$
f	f	f	f	f
$\top$	$\top$	f	$\top$	$\perp$
$\perp$	$\perp$	f	$\perp$	$\perp$

The truth table for negation is very simple. The interpretation of $\perp$ was: neither a statement nor its negation has been told. Therefore, if A is underdetermined, so is $\neg A$. A similiar argument holds for $\top$.

	t	f	$\top$	$\perp$
$\neg$	f	t	$\top$	$\perp$

The definition of $A \vee B$ as $\neg(\neg A \wedge \neg B)$ and $A \rightarrow B$ as $\neg A \vee B$ yields

$\vee$	t	f	$\top$	$\perp$
t	t	t	t	t
f	t	f	$\top$	$\perp$
$\top$	t	$\top$	$\top$	$\perp$
$\perp$	t	$\perp$	$\perp$	$\perp$

$\rightarrow$	t	f	$\top$	$\perp$
t	t	f	$\top$	$\perp$
f	t	t	t	t
$\top$	t	$\top$	$\top$	$\perp$
$\perp$	t	$\perp$	$\perp$	$\perp$

The above tables show that $A \rightarrow A$ is no theorem[6] and more than that: our logic (if we call the set of truth tables a logic) does not have any tautologies. Indeed, this is very bizarre. A closer look shows that there is no well-formed formula to represent the epistemic state of A is underdetermined, but there are wffs which represent the other truth values. For example 'A has been told to be true' is represented by $\{A\}$, wheras 'A has been told both, to be true and to be false' can be represented by $\{A, \neg A\}$. If we choose the empty set to represent A is underdetermined, we cannot distinguish between A is underdetermined and B is underdetermined. Therefore, we introduce a new operator, ∇ (which is similiar to the T-supplement in three-valued logic) to denote A is underdetermined by ∇A, associating with ∇ the following truth table:

	t	f	$\top$	$\perp$
∇	f	f	f	t

Let t_4 be the tupel $[(\rightarrow, \neg, \nabla), \{t, f, \top, \perp\}, \{t, \top\}]$.

2.2 Truth functional completeness

Are the truth tables for $\neg, \rightarrow, \nabla$ given in section 2.1 sufficient to generate all possible truth tables, i.e., are they functionally complete?

It is easy to see that t_4 is not functionally complete. For example consider the function $g : \tau \rightarrow \tau$ such that $g(t) = g(f) = g(\top) = g(\perp) = \top$. Clearly, g cannot be defined using $\rightarrow, \neg$ and ∇. The reason is that there a subset $X \subset \tau$, namely $\{t, f\}$, which is closed under $\rightarrow, \neg$ and ∇. One way to solve the problem is to define a truth function (shift negation[7], denoted by $\triangleleft$) in order to obtain truth functional completeness. The following will do:

	t	f	$\top$	$\perp$
$\triangleleft$	$\perp$	t	f	$\top$

In this section we replace the ∇-operator by $\triangleleft$; let $T_4 := [(\rightarrow, \neg, \triangleleft), \{t, f, \top, \perp\}, \{t, \top\}]$. A class of truth value functions G is called *functionally complete* if and only if every other truth value function can be expressed in terms of the functions in G. A function is T_4-definable iff it can be expressed in terms of $\neg, \rightarrow$ and $\triangleleft$. Furthermore let F be a class of functions defined on $\tau = \{t, f, \perp, \top\}$; a subset $X \subseteq \tau$ is said to be *F-closed* if X is closed under the application of all $f \in F$. The *F-closure* of $X \subseteq \tau$ is the smallest F-closed set containing X.

[6] The fact that $A \rightarrow A$ is no theorem is worse than one would presume; $\not\models A \rightarrow A$ means that the normalization condition (NC) for implication (i.e. $A \rightarrow B$ takes a non-designated truth value if and only if $I(A)$ is a designated one but not $I(B)$) is not fulfilled. This in turn means that the Rosser/Turquette procedure for axiomatizing the set of tautologies cannot be applied. Additionally NC is also violated by the fact that $A \rightarrow B$ takes the truth value $\top$ for $I(A) = \top$ and $I(B) = f$; but this in turn guarantees the formal treatment of MOBAL's blocking behaviour or more generally: a nonmonotonic behaviour of the entailment relation of section 3 which is based on T_4 . Indubitably one could reasonably ask whether there ain't no other pssobility to simulate blocking without the infringement of NC. The answer is 'No, not within a truth-functional (extensional) semantics'. This can be seen by looking at the basic assumption of nonmonotonic systems; $A, A \rightarrow B$ should be valid but B could not be satisfied.

[7] After finishing the paper we found out that an identical function ($\sim_m$) has been found by Post in 1921 ([Post, 1921]).

In order to prove the functional completeness of T_4 , we first show that the functions $\Phi_k : \tau \to \tau$ are T_4 -definable[8].

$$\Phi_k(x) = \begin{cases} t & \text{for } x = k \\ f & \text{otherwise} \end{cases}$$

Lemma 1 *$\Phi_k(x)$, $k \in \tau$ is T_4 -definable.*

All proofs missing in the body of the paper can found in the Appendix. The following lemma has been adopted from [Urquhart, 1986] (Lemma 2.9)

Lemma 2 *Let F be a set of functions on $\tau = \{t, f, \top, \bot\}$ containing $\neg, \to, \triangleleft$ as well as all the Φ_k. Then an n-place function g is F-definable if and only if $g(\vec{x})$ is in the F-closure of $\{x_1, \ldots, x_n\}$, for all $\vec{x}$ in τ.*

Theorem 1 *T_4 is functionally complete.*

Proof We will show that every unary function is T_4 definable and that the only non-empty T_4 -closed subset of τ is τ itself. Hence, T_4 is functionally complete by Lemma 2. The function $T(x) = t$ for all x is T_4 -definable:

$$T(x) = \neg(x \to x) \to (\triangleleft x \to \neg \triangleleft \triangleleft x)$$

Furthermore the functions T_k for any $k \in \tau$ are defined as

$$T_k(x) := \begin{cases} f & \text{for } x \neq t \\ t & \text{otherwise} \end{cases}$$

are T_4 -definable as proved by the following:

$$\begin{aligned} T_f(x) &= \neg T(x) \\ T_t(x) &= \Phi_t(x) \\ T_\bot(x) &= \neg \triangleleft \Phi_t(x) \\ T_\top(x) &= \neg \triangleleft \triangleleft \Phi_t(x) \end{aligned}$$

Now let $g(x)$ be any one place τ-valued function. Then the following function coincides with g:

$$T_{g(t)}(x) \vee T_{g(f)}(\triangleleft x) \vee T_{g(\top)}(\triangleleft \triangleleft x) \vee T_{g(\bot)}(\triangleleft \triangleleft \triangleleft x)$$

The application of Lemma 2 yields the functional completeness of T_4 .

◀

As a general remark on functional completeness of T_4 let us state that although there is no intuitive reading of the one place function $\triangleleft$, it provides the basis of functions that make more sense, e.g., ∇.

[8] Don't try to look for a natural meaning of these functions; we don't have any in mind. As in [Urquhart, 1986] they serve as mathematical construction necessary for functional completeness.

2.3 Relationship between T_4 and Kleene's K_3

This short section relates the above truth tables to the well known strong three-valued logic K_3. As we will discuss in section 4.1 the philosophical implications are far-reaching. This is due to the fact that 'overdefinedness' and 'underdefinedness' have the same epistemological status; both lack information, although the set of valid sentences increases[9]. We took this into account by the definition of the truth tables: both truth values $\top$ and $\perp$ could change to t or f in a future epistemic state. These issues will be treated in a later section. We restrict ourselves in this section to the meaning of the basic connectives.

Within three-valued logics there are different interpretations of the third truth value I. One is to read I as 'undefined', another one is 'meaningless' and a third one is 'paradoxical'[10] There is a close relationship between Kleene's strong three-valued logic K_3 and the system T_4 . To see this, we restrict our truth tables to three truth values: t, f, X where X may be substituted by $\top$ or $\perp$.

Consider the following truth tables:

$\vee$	t	X	f
t	t	t	t
X	t	X	X
f	t	X	f

$\rightarrow$	t	X	f
t	t	X	f
X	t	X	X
f	t	t	t

$\wedge$	t	X	f
t	t	X	f
X	X	X	f
f	f	f	f

	t	X	f
$\neg$	f	X	t

The reader may check that these tables correspond to those of section 2.1. Furthermore these truth tables are exactly those defined by Kleene's strong three-valued logic K_3 ([Kleene, 1952]). The only but important difference is that K_3 has only t as designated truth value. Of course, since we did not compare Kleene's notion of entailment with ours, we cannot say that our logic is really a combination of Kleene's *logic*, but as far as the truth tables are concerned, it is.

3 Entailment Relations

In this section we will consider the propositional case of our four-valued logic. That is, we assume the classical propositional language $\mathcal{L}_c$ with the connectives of T_4 . First, we will develop a notion of semantic entailment before we investigate a syntactic characterization of this consequence relation. In the following *APROP* and *PROP* denote the sets of atomic propositions and propositions, respectively.

Let I be an interpretation function which maps every sentence of a propositional language into the set $\tau = \{t, f, \top, \perp\}$ of truth values. Now let $D \subseteq \tau$ be the set of designated truth values. We say that I satisfies an atomic proposition P (denoted by $I \models P$), if $I(P) \in D$. The satisfiability relation $\models$ can be extended as usual in order to capture compound propositions. If X is a set of propositions we say that $I \models X$ if I satisfies every element of X.

[9] Unlike Belnap's four-valued logic where knowledge increases with every new sentence, T_4 treats overdetermined formulas as uncertain (i.e., knowledge decreases).

[10] Note that we have different truth values namely $\top$ and $\perp$ to denote 'undefined' and 'paradoxical'.

A sentence A is a consequence of a set X of sentences (or: X entails A, denoted by $X \Vdash A$) if every model of X is also a model of A. This is the standard Bolzano definition of entailment, which unfortunately does not meet our aims. Consider the set $X = \{A, A \to B\}$. We wish that this set entails B. But there is an interpretation $I(A) = \top, I(B) = f$ which satisfies X but not B. The reason is that there are two designated truth values: t and $\top$; if we choose $I(A) = \top$, we would satisfy $\neg A$ as well. But $\neg A$ has not been mentioned in our set of formulas. What we've got here is a kind of over-interpretation. We have to look for a kind of 'minimal truth assignment'. Therefore we assume an ordering relation $\leq$ on the set of truth values. In our case we define $\perp \leq t \leq \top$ and $\perp \leq f \leq \top$[11].

Definition 1 ($\models$, $\dot{\models}$) *Let $I : APROP \to \tau$ be an interpretation, D the set of designated truth values and A a formula. We define a relation $\models$ between an interpretation and a formula*

$$I \models A \textit{ if } I(A) \in D$$

and relation $\dot{\models}$ between an interpretation and a set of propositions

$$I \dot{\models} X \textit{ if } I \models A \textit{ for all } A \in X \textit{ and } I \textit{ is minimal w.r.t } \leq .$$

The sets MOD *and* $\mathrm{M\overset{\bullet}{O}D}$ *can be defined as usual:* $\mathrm{MOD}(X) := \{M | M \models X\}$ *and* $\mathrm{M\overset{\bullet}{O}D}(X) := \{M | M \dot{\models} X\}$.

We say that a formula A is *satisfiable (valid)* if there is (for all) an interpretation I such that $I \models A$. It is now easy to see that $I(A) = \top, I(B) = f$ is no minimal model for $X = \{A, A \to B\}$, since there is an interpretation $I'(A) = t$ and $I'(B) = t$ which satisfies X, and for every atomic proposition p, $I'(p) \leq I(p)$.

The relation $\dot{\models}$ is useful for defining an appropriate entailment relation $\Vdash$:

Definition 2 ($\Vdash$, **ENT**) *Let X be a set of formulas, A a formula. We say that $X \Vdash A$ if for every I such that $I \dot{\models} X$, $I \models A$. The operator* ENT *is defined as* $\mathrm{ENT}(X) := \{A | X \Vdash A\}$.

Let us call $\mathcal{L}_4 = \langle T_4, \Vdash, \dot{\models} \rangle$.

Proposition 1

1. $X \subseteq \mathrm{ENT}(X)$ *(Inclusion)*
2. *If $X \Vdash A$ then for every $M \dot{\models} X$, $M \dot{\models} A$ (Dot entailment)*
3. *If $X \Vdash A$ and $X \Vdash B$ then $X \cup \{A\} \Vdash B$ (Cumulativity)*
4. *$X \Vdash A$ does not imply that $X \cup \{B\} \Vdash A$ for an arbitrary B (Nonmonotonicity)*
5. $\mathrm{ENT}(X) = \mathrm{ENT}(\mathrm{ENT}(X))$ *(Idempotency)*

Corollar 1 *If* $\mathrm{M\overset{\bullet}{O}D}(X) \subseteq \mathrm{M\overset{\bullet}{O}D}(A)$ *then* $\mathrm{M\overset{\bullet}{O}D}(X) \subseteq \mathrm{M\overset{\bullet}{O}D}(X \cup \{A\})$

[11] This is exactly Belnap's lattice **A4**.

The next observation is closely related to Section 1.1, where we pointed out that either 'inconsistencies are inferred' or 'from someting inconsistent nothing should be inferred' is a plausible scheme of reasoning. In our logic inconsistencies are allowed in the sense that they don't produce the set of all wffs. If a set of formulas T is consistent in a classical sense of the term (i.e. $A \wedge \neg A \notin T$), so is the $\Vdash$-closure.

Observation 1 *Let T be a set of wffs and A a formula. $\{A, \neg A\} \nsubseteq T$ if and only if $T \nVdash A \wedge \neg A$.*

We now come to some brief observations, which characterize the relationship between our four-valued semantics and the classical two-valued semantics. In order to describe the relationship we restrict the set of well-formed formulas of $\mathcal{L}_4$ to the language $\mathcal{L}_c$ of classical propositional calculus. In the following we denote by I_4 (I_2) a four-valued (two-valued) interpretation of the sentences in $\mathcal{L}_c$ and by $\nVdash_c$ the classical entailment relation. The concepts of satisfiability and validity for the two-valued semantics are defined as usual.

The first observation is trivial:

Observation 2 *Let $X \subseteq \mathcal{L}_c$ be a set of formulas. If X has a two-valued model, then there is a four-valued model for X.*

Observation 3 *Let $\mathcal{L}_c$ be the classical propositional language, X a set of formulas. For every formula A, which is satisfiable, but not valid concerning X, the following holds:*

1. *$X \nVdash_c A$*
2. *$X \Vdash \nabla A$ and $X \Vdash \nabla \neg A$ in our extended language iff there is a unique I such that $I \models^{\bullet} X$*

4 Axiomatizability and other problems

In classical logic - and this is easy to see - the axiomatizability of the entailment relation can be guaranteed by the facts that the set of tautologies is axiomatizable, the entailment relation is monotonic and modus ponens (MP) is a correct rule of proof. For $\mathcal{L}_4$, however, MP is not correct for arbitrary sets of formulas, i.e., no correct rule of inference.

Observation 4 *If $A, A \rightarrow B$ are tautologies, then MP is a correct of proof, i.e., B is also a tautology.*

4.1 Belief revision and reasoning by defaults

It has been mentioned by Makinson and Gärdenfors that belief-revision and nonmonotonic reasoning are two sides of the same coin and Makinson published a method for transferring one system into the other. Despite these results we think that the motivation for using a logic for nonmonotonic reasoning differs from that of belief revision. A logic for non-monotonic reasoning tries to capture non-deductive reasoning in a deductive formalism (reasoning based on assumptions). In our terms 'reasoning with assumptions' means that there is a normative reassignment of truth values; formulas which had the truth value $\perp$

(or in classical terms, which were not valid but only satisfiable) get one of the truth values t or f. Once they have this interpretation, it will never become uncertain again. A logic for nonmonotonic reasoning only shifts a formula's assignment from the bottom level to the mid level of the truth value lattice (shown in Figure 4.1); once things are certain they will remain certain. On the contrary belief revision cannot guarantee this property. If a set of beliefs has been revised (or better contracted) nobody will witness that the contracted formula will not be added at a later stage. Ergo, the nonmonotonicity of belief revision is 'not as safe as' than that of a typical logic for reasoning based on assumptions. Within the first one a sentence or a belief may fall from the top to the middle of the lattice; new information may jostle it to the top.

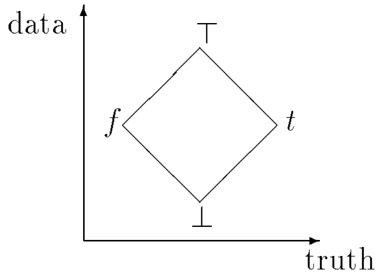

Figure 4.1 Belnap's lattice **A4**.

4.2 Conclusions

Starting from a wishlist for a notion of paraconsistent entailment, we developed a four-valued semantics for the connectives $\neg$ and $\wedge$. This semantics turned out to be a combination of two variations on Kleene's strong three-valued logic K_3 which means that the truth values $\top$ and $\bot$ in our four-valued semantics correspond to the notion of a truth value gap in K_3. This in turn implies that the epistemic status of an overdetermined and underdetermined sentence in our semantics is the same: both represent a lack of knowledge, albeit the amount of told sentences (data) differs. Coherently, the entailment relation which is based on a notion of minimal models does not allow the entailment of a sentence via disjunctive syllogism with overdetermined premisses. This type of paraconsistent entailment is very cautious and maybe the least possible one in the sense that it isolates classical inconsistencies. Table 4.2 illustrates the major differences between $\mathcal{L}_4$, Belnap's four-valued logic **A4** and MOBAL's inferential behaviour:

Formulas	MOBAL	**A4**	$\mathcal{L}_4$
$X = \{A, \neg A, A \rightarrow B\}$	$X \not\vdash_M B$	$X \Vdash_B B$	$X \not\Vdash B$
$X = \{A, \neg B, A \rightarrow B\}$	$X \vdash_M B$	$X \Vdash_B B$	$X \not\Vdash B$

Table 4.2

$\vdash_M$ denotes MOBAL's inference relation and $\Vdash_B$ denotes Belnap's entailment relation.

The next step is to axiomatize the set of all tautologies via a sequent-style calculus,i.e., to develop a proof theory for $\mathcal{L}_4$.

Acknowledgements We would like to thank Joachim Hertzberg and Anke Rieger for fruitful discussions and enlightening comments. This work was partly supported by the European Community (ESPRIT Basic Research Action 6020, Inductive Logic Programming).

5 Appendix: Proofs

Lemma 1 *$\Phi_k(x)$, $k \in \tau$ is T_4 -definable.*

Proof We will define the corresponding functions Φ_k for every $k \in \tau$. Define

$$g(x) = \begin{cases} t & \text{for } x = t \text{ or } x = f \\ f & \text{for } x = \top \\ \top & \text{for } x = \bot \end{cases}$$

$g(x) = \triangleleft\neg(x \rightarrow x)$ is L_4-definable. Now consider the following function:

$$f(x) = \begin{cases} t & \text{for } x \in \{t, f, \top\} \\ \bot & \text{otherwise} \end{cases}$$

$f(x) = \neg(x \rightarrow x) \rightarrow (\triangleleft x \rightarrow \triangleleft x)$ is L_4-definable. The reader may verify that the following equations hold:

$$\begin{aligned} \Phi_\bot(x) &= \neg(g(g(f(x)))) && (1) \\ \Phi_t(x) &= \Phi_\bot(\triangleleft x) && (2) \\ \Phi_f(x) &= \Phi_\bot(\triangleleft \triangleleft x) && (3) \\ \Phi_\top(x) &= \Phi_\bot(\triangleleft \triangleleft \triangleleft x) && (4) \end{aligned}$$

Note that $\Phi_\bot$ is exactly the same as the ∇-operator.

◀

Lemma 2 *Let F be a set of functions on $\tau = \{t, f, \top, \bot\}$ containing $\neg, \rightarrow, \triangleleft$ as well as all the Φ_k. Then an n-place function g is F-definable if and only iff $g(\vec{x})$ is in the F-closure of $\{x_1, \ldots, x_n\}$, for all $\vec{x}$ in τ.*

Proof For the $\Rightarrow$ part assume that g is F-definable. By definition of the F-closure of $\{x_1, \ldots, x_n\}$ it follows that $g(\vec{x})$ is in the F-closure of $\{x_1, \ldots, x_n\}$. For the converse ($\Leftarrow$-part) assume that $g(\vec{x})$ is in the F-closure of $\{x_1, \ldots, x_n\}$. Now consider the function Ψ as defined below and an n-tupel $\vec{a}$ in τ.

$$\Psi(\vec{a}) := \neg\Phi_{a_1}(x_1) \vee \ldots \vee \neg\Phi_{a_n}(x_n) \vee g(\vec{a})$$

Clearly, $\Psi(\vec{a})$ takes the value $g(\vec{a})$ if $\vec{x} = \vec{a}$ and t otherwise. Now let $\vec{a}_1, \ldots, \vec{a}_{n!}$ be all possible disjunct inputs for Ψ. Then

$$g(\vec{x}) = \Psi(\vec{a}_1) \to \Psi(\vec{a}_2) \to \ldots \to \Psi(\vec{a}_{n!})$$

Hence, g is F-definable.

◀

Proposition 1 *1.* $X \subseteq \mathrm{ENT}(X)$ *(Inclusion)*

2. If $X \Vdash A$ *then for every* $M \overset{\bullet}{\models} X$, $M \overset{\bullet}{\models} A$ *(Dot entailment)*

3. If $X \Vdash A$ *and* $X \Vdash B$ *then* $X \cup \{A\} \Vdash B$ *(Cumulativity)*

4. $X \Vdash A$ *does not imply that* $X \cup \{B\} \Vdash A$ *for an arbitrary* B *(Nonmonotonicity)*

5. $\mathrm{ENT}(X) = \mathrm{ENT}(\mathrm{ENT}(X))$ *(Idempotency)*

Proof

Inclusion Trivial. Since from $M \overset{\bullet}{\models} \{A\}$ it follows that $M \overset{\bullet}{\models} \{A\}$, we have $\{A\} \Vdash A$, the relation $\Vdash$ is reflexive.

Dot entailment Trivial. Since M is an interpretation function which assigns each atom a truth value and the function does not change.

Cumulativity We know that $M\overset{\bullet}{O}D(X) \subseteq M\overset{\bullet}{O}D(A)$ and $M\overset{\bullet}{O}D(X) \subseteq M\overset{\bullet}{O}D(B)$. The proposition follows if we can show that if $X \Vdash A$ then $M\overset{\bullet}{O}D(X) = M\overset{\bullet}{O}D(X \cup \{A\})$. We show this by contradiction. Therefore assume that $M\overset{\bullet}{O}D(X) \neq M\overset{\bullet}{O}D(X \cup \{A\})$. Then there are two cases:

1. Assume there is an M such that $M \overset{\bullet}{\models} X$ and $M \not\overset{\bullet}{\models} X \cup \{A\}$. This means that the conjunction of all elements of the set $X \cup \{A\}$ is not entailed by X. Since $\Vdash$ is refelxive, $X \nVdash A$, which is a contradiction.
2. Assume there is $M \overset{\bullet}{\models} X \cup \{A\}$ but $M \not\overset{\bullet}{\models} X$. Hence, $X \cup \{A\} \nVdash X$ in contradiction to the inclusion property.

Nonmonotonicity Trivial, by counter-example.

Idempotency The relation $ENT(X) \subseteq ENT(ENT(X))$ follows from the inclusion. It remains to show that $ENT(ENT(X)) \subseteq ENT(X)$. We have to show that for every

$$A \in ENT(ENT(X)) \text{ it follows that } A \in ENT(X) \tag{5}$$

Define

$$\Phi := ENT(X) - X$$

Obviously, $X \Vdash \Phi$. A reformulation of 5 using Φ

$$X \cup \Phi \Vdash A \text{ implies } X \Vdash A$$

Thus we have to show that

$$\overset{\bullet}{MOD}(X \cup \Phi) \subseteq \overset{\bullet}{MOD}(A) \text{ and } \overset{\bullet}{MOD}(X) \subseteq \overset{\bullet}{MOD}(\Phi) \text{ implies } \overset{\bullet}{MOD}(X) \subseteq \overset{\bullet}{MOD}(A)$$

Since $\overset{\bullet}{MOD}(X) \subseteq \overset{\bullet}{MOD}(\Phi)$ we have by Corollar 1 that $\overset{\bullet}{MOD}(X) \subseteq \overset{\bullet}{MOD}(X \cup \Phi)$ and from $\overset{\bullet}{MOD}(X \cup \Phi) \subseteq \overset{\bullet}{MOD}(X)$ it follows that

$$\overset{\bullet}{MOD}(X) \subseteq \overset{\bullet}{MOD}(A)$$

◀

Corollar 1 *If* $\overset{\bullet}{\mathrm{MOD}}(X) \subseteq \overset{\bullet}{\mathrm{MOD}}(A)$ *then* $\overset{\bullet}{\mathrm{MOD}}(X) \subseteq \overset{\bullet}{\mathrm{MOD}}(X \cup \{A\})$

Proof Obviously: for all $M \in \overset{\bullet}{MOD}(X) : M \models A$. Hence $M \models X_1 \ldots \wedge \ldots X_n \wedge A$ $(X_1, \ldots, X_n \in X)$.

◀

Observation 3 *Let $\mathcal{L}_c$ be the classical propositional language, X a set of formulas. For every formula A, which is satisfiable, but not valid concerning X, the following holds:*

1. $X \nVdash_c A$

2. *$X \Vdash \nabla A$ and $X \Vdash \nabla\neg A$ in our extended language iff there is a unique I such that $I \overset{\bullet}{\models} X$*

Proof

1. Obvious, by definition.

2. We have to proof that $I_4(A) = \perp$ is the only allowed interpretation. From the satisfiablity of A w.r.t X we conclude that $I_2(A) = t$ and $I_2'(A) = f$ are possible interpretations within the two-valued semantics. But the greatest lower bound of t and f is according to our lattice $\perp$ which is also a possible interpretation. Moreover $I_4(A) = \perp$ is the only possible interpretation for $\overset{\bullet}{\models}$. Hence, $X \Vdash \nabla A$ and $X \Vdash \nabla\neg A$.

◀

References

[Gabbay, 1982] Dov M. Gabbay *Intuitionistic Basis for Non-Monotonic Logic*, PROCEEDINGS OF THE 6TH CONFERENCE ON AUTOMATED DEDUCTION, LNCS 138, Springer, Berlin, 1982.

[Kleene, 1952] S.C. Kleene *Introduction to Metamathematics*, Van Nostrand, New York, 1952.

[Morik et al.,1933] K. Morik, S. Wrobel, J. Kietz and W. Emde *Knowledge Acquisition and Machine Learning*, Academic Press, London, 1933.

[Servi, 1992] Gisèle Servi *Nonmonotonic Consequence Based on Intuitionistic Logic*, The Journal of Symbolic Logic, Vol. 57(4), 1992.

[Urquhart, 1986] A. Urquhart *Many-Valued Logic* in *Gabbay and Guenthner (Eds.)* Handbook of Philosophical Logic, Vol. III, Reidel, Dordrecht, 1986.

[Post, 1921] E.Post *Introduction to a general theory of elementary propositions*, American J. o. Math., Vol. 43, 163-185, 1921.

[Belnap, 1977] N.Belnap *A useful four-valued logic*, in Modern uses of Multiple-valued logic, edited by J.M.Dunn and G. Epstein, Reidel, Dordrecht, 1977.

[Rescher, 1969] N. Rescher *Many-valued Logic*, McGraw Hill, New York, 1969.