PRUNING MULTI-EXTENSIONS VIA EXCEPTIONS[1]

Marcelino Pequeno[2]
Laboratory of Artificial Intelligence (LIA)
Department of Computing
Federal University of Ceara — Brazil
E-mail: marcel@lia.ufc.br

1 Introduction

The main contribution of this paper is a method for pruning multi-extensions of a defeasible theory by using the exceptions to order the defeasible formulae.

We construct a defeasible logic — *DEFEASIBLE LOGIC WITH EXCEPTIONS FIRST (DLEF)* — in which extensions are built taking into account the order on the defeasible formulae induced by the exceptions.

This device prompts DLEF as a powerful tool to formalize common sense reasoning. It is on the formalization of the frame problem that we best evaluate the original features of DLEF. DLEF allows the formalization of the *persistence axiom* in the temporal projection problem in a stepwise way. That is, the persistence axiom is applied *locally* after every action is performed. Thus, if no exception to some properties is present while an action is performed the persistence axiom is used to conclude that those properties will remain unaltered in the resulting situation. Therefore, no property at the present is changed just for the sake of not changing some other properties in the future. The only reason for changes in properties are explicit changes provoked by the action being performed at the moment.

It is straightforward to see that with this formalization of the persistence axiom DLEF precludes in a general and natural way unwanted extensions on temporal projection problems including the now (in)famous Yale Shooting Problem [H&M87].

The rest of this paper is structured as follows: In section 2, we present the syntax of DLEF. Section 3 presents the order on the defeasible formulae induced by the exceptions. In section 4, we build DLEF extensions for defeasible theories. The stepwise approach to the frame problem is presented in section 5. Finally, section 6 presents our conclusions and further development to DLEF.

---

[1] Paper presented in the II Dutch/German Workshop on Non-Monotonic Reasoning, Utrecht, March, 29 - April, 1, 1995.

[2] On leave from State University of Ceara.

2 Syntax of DLEF

The ontology of DLEF comprises three kinds of expressions. Absolute, conclusive formulae or *facts;* defeasible, inconclusive formulae or *defaults;* and exceptions or *defeaters.*

The language L_A of the absolute and defeasible formulae consists of all ordinary first order formulae (wff's) which can be formed using an alphabet A consisting of countably many variables $x, y, z, x_1, y_1, z_1, \ldots$, countably many functions letters $a, b, c, f, g, h, a_1, b_1, c_1, \ldots$, countably many predicate letters $P, Q, R, P_1, Q_1, R_1, \ldots$, the usual punctuation signs ',', '(' and ')', and the standard logical constants $\neg$ (not), $\wedge$ (and), $\vee$ (or), $\Rightarrow$ (implies) and quantifiers $\forall x$ (for all x), $\exists x$ (there exists an x). 0-ary function letters will sometimes be called constants. Terms, atoms and literals are defined in the standard way [End72]. To represent defeaters a new connective '//' is introduced. A defeater is any expression of the form

(1) $\alpha(\bar{x}) // \omega(\bar{x})$,

where $\alpha, w \in L_A$ are wff's whose free variables are among those of $\bar{x} = x_1, \ldots, x_n$. $\alpha(\bar{x})$ is called a *defeater* or *exception* for $w(\bar{x})$, and $w(\bar{x})$ is called the ***head*** of the defeater. Expression (1) intuitively reads as *"if something is an* α *then the defeasible formula w is not applicable for it"*, *i.e.*, α blocks the application of w.

An *axiomatic base,* or a *defeasible theory* is a triple <W, D, U>, where W is a set of closed wff's of L_A, D is a set of (possibly open) wff's of L_A, and U is a set of (possibly open) defeaters for the wff's in D. W and D are intended to represent the conclusive and inconclusive statements respectively, and U is intended to represent the exceptions to the statements in D.

The open expressions in D and U are meant as schemata representing the set of all ground instances of the expressions over the language L_A.

A defeasible theory <W, D, U> is *closed* iff every defeasible formula of D and every defeater of U are closed.

The language of DLEF allows a flexible representation of the defeasible information. Specially, as defeaters are written separately from the defeasible formulae, it allows a modular representation of the exceptions to the defeasible formulae. New exceptions are added by just writing new defeaters with no need to change any formula already in the theory.

3 Partial Order on the Defeasible Formulae

Given a closed defeasible theory T = <W, D, U> we order the defeasible formulae in *D* in such a way that defeasible formulae that *support* an exception for a defeasible formula have higher priority than the latter on the construction of the extension. Thus if the *supporting* formulae are compatible with the rest of the formulae in the extension, the exception will be carried into the extension. Consequently, the respective defeasible formula will be drawn off the extension.

Next we define precisely the order on the defeasible formulae in a defeasible theory. For the definitions below let T = <W, D, U> be a closed defeasible theory.

Definition 1 A set $R \subseteq D$ *supports* a sentence α of L_A with respect to the closed defeasible theory T = <W, D, U>, iff (i), (ii) and (iii) hold. Notice that 'Ã' stands for first order logic derivation.

(i) $W \cup R$ is first order logic consistent;

(ii) $W \cup R$ Ã α;

(iii) *R* is *minimal, i.e.,* if $W \cup R'$ is consistent, $W \cup R$ Ã α and $R' \subseteq R$, then $R' = R$.

R is said to be a *supporting set* for α with respect to the theory *T*. We also say that every member *d* of *R supports* α or that *d* is a *supporter* for α, and we represent it as $d \ll \alpha$.

Observe that if *W* is inconsistent then the support relation is empty.

Condition (iii) guarantees that if *d* is a supporter for α, then *d* is *relevant* to a first order derivation of α.

Definition 2 A defeasible formula d_1 is *immediately smaller* than another defeasible formula d_2 in a closed defeasible theory T = <W, D, U>, denoted by $d_1 \langle d_2$, iff there exists a defeater $\alpha // d_2 \in U$ and d_1 supports α ($d_1 \ll \alpha$).

Definition 3 The partial order on the defeasible formulae of a closed defeasible theory T = <W, D, U>, '$\angle$', is defined as the transitive closure of the relation '$\langle$', *i.e.,* the following conditions hold.

(i) if $d_1 \langle d_2$, then $d_1 \angle d_2$.

(ii) If $d_1 \angle d_2$, and $d_2 \langle d_3$ then $d_1 \angle d_3$.

The order on the defeasible formulae plays a fundamental role on the definition of extension in the next section (condition (3) in definition 7). First, it allows a constructive, though not effective, characterization of an extension (theorem 1). Secondly, it enables DLEF to preclude some extensions. Let *d* be a ground defeasible formula and $\alpha // d$ be a defeater, that is, α is an exception to d in the theory in hand. Since all supporters for α are smaller than d, they will have preference over d on the construction of the extension. Thus α will be put into the extension (provided its supporters are compatible with the other formulae in the extension) before than d, in consequence d will be ruled out. This is how DLEF prunes the set of extensions for a defeasible theory.

Example 1 Let T = <W, D, U> be a defeasible theory where

$$W = \{A, C, B \Rightarrow A\}$$
$$D = \{A \Rightarrow \neg F, C \Rightarrow B, B \Rightarrow F\}$$
$$U = \{B \,//\, (A \Rightarrow \neg F)\}$$

In this theory $C \Rightarrow B \angle A \Rightarrow \neg F$ (note that $C \Rightarrow B \ll B$ e $B \,//\, A \Rightarrow \neg F$). We shall show in next section that this theory has only one extension in DLEF, namely $E = Th(W \cup \{C \Rightarrow B, B \Rightarrow F\})$. For an intuitive reading of the example, let A stands for *Animal; B* for *Bird; C* for *Canary* and *F* for *Flies.*

[End_of_Example]

Definition 4 A closed theory T = <W, D, U> is *ordered* or *acyclic* iff for no formula $d \in D$, $d \angle d$.

Example 2 Let T = <W, D, U> be a defeasible theory where

$$W = \varnothing$$
$$D = \{B, C\}$$
$$U = \{C \,//\, B, B \,//\, C\}$$

This theory is not ordered since $B \angle C$ and $C \angle B$, hence $B \angle B$.

[End_of_Example]

Etherington [Eth88] proposes an order on default theories that resembles very much the order introduced here on defeasible theories. In [Peq94], we show that a defeasible theory is ordered if its corresponding default theory is ordered according to Etherington's definition, but the converse is not valid. Etherington shows that ordered default theories are *coherent,* in the sense that they have extensions. In DLEF, by definition, only ordered theories have extensions, and all ordered defeasible theories have extensions as we prove in [Peq94] (notice that theorem 1 is a constructive way of calculating an extension for DLEF).

4 Building Extensions

Although a defeasible theory T = <W, D, U> might contain some open defeasible wff's and defeaters, we shall always associate to it a closed defeasible theory called the associated ground defeasible theory, consisted of *W* (all formulae of *W* are closed) and the set of all ground instances (over the Herbrand universe of the language of *T*) of the expressions in *D* and *U*, respectively. We only define extensions for closed theories, extensions to open theories are determined assuming they represent the associated ground theories.

The main distinction from DLEF to other nonmonotonic logics lies in the fact that DLEF takes into account the order on the defeasible formulae defined in section 3 to prune the set of extensions for a defeasible theory.

Some preliminary concepts are necessary before defining extensions for defeasible theories. In the following definitions and lemmas let T = <W, D, U> be a closed ordered defeasible theory, and '$\angle$' be the order on the defeasible formulae associated to *T*.

Definition 5 A defeasible formula $d \in D$ is *applicable* in a set *S* of formulae with respect to *T* iff:

(a) *d* is consistent with *S*, and

(b) For all $(\alpha // d) \in U$, $\alpha \notin Th(S)$.

Definition 6 A defeasible formula *d* is *reachable* in a set of formulae *S* with respect to *T*, iff for all defeasible formula $d' \in D$ such that $d' \angle d$, either $d' \in Th(S)$ or *d'* is not applicable in *S*.

Extensions are only defined for closed ordered defeasible theories. Unordered theories have no extensions in DLEF.

Definition 7 Let T = <W, D, U> be a closed ordered defeasible theory.
$E = Th(W \cup S)$, $S \subseteq D$, is an extension for *T* iff:

1. *E* is *maximal, i.e.*, if $d \in D$ and *d* is reachable and applicable in *E*, then $d \in S$.
2. *E* is *sound, i.e.*, if $d \in S$ then *d* is reachable and applicable in *E*.
3. *E* complies with the *exceptions-first* principle, *i.e.*, for all $d \in S$, if $d' \angle d$ and $d' \notin S$, then d' is not applicable in $Th(W \cup \not{S}_{d'}\}$, where for all $d \in S$ $\not{S}_d = S - S_{>d}$, and $S_{>d} = \{\alpha \in S; d \angle \alpha\}$.

The set *S* is called a *generating set* for the extension *E* with respect to *T*.

We can find a constructive though not effective characterization of extensions for defeasible theories.

Theorem 1 Let T = <W, D, U> be a closed ordered defeasible theory. Define

$S_0 = \varnothing$

$E_0 = W$

And for $i \geq 0$,

$E_{i+1} = Th(E_i) \cup S_{i+1}$, where $S_{i+1} \subseteq D$ and

a) $E_i \cup S_{i+1}$ is consistent;

b) If $d \in S_{i+1}$ then d is reachable and applicable in E_i.

(The idea is that S_{i+1} extends E_i with defeasible formulae from D which are reachable and applicable in E_i)

Let $S = \bigcup_{i=0}^{\infty} S_i$.

If $E = \bigcup_{i=0}^{\infty} E_i = Th(W \cup S)$ and S is *maximal,* in the sense that if $d \in D$ is applicable and reachable in E then $d \in S$, then E is an extension for T.

This theorem is proved in [Peq94]. Perhaps the best way to grasp the feeling of our definition is seeing an example. In the example below let T = <W, D, U> be a defeasible theory.

Example 1 revisited.

$W = \{A, C, B \Rightarrow A\}$

$D = \{A \Rightarrow \neg F, C \Rightarrow B, B \Rightarrow F\}$

$U = \{B \ // \ (A \Rightarrow \neg F)\}$

This theory has only one extension:

$E = Th(W \cup \{C \Rightarrow B, B \Rightarrow F\})$.

Notice that $(C \Rightarrow B) \angle (A \Rightarrow \neg F)$.

Observe that $E' = Th(W \cup \{A \Rightarrow \neg F, B \Rightarrow F\})$ is not a DLEF extension for T.

[End_of_Example]

5 The Frame Problem

DLEF allows the formalization of the *persistence axiom* in the temporal projection problem in a stepwise way. That is, the persistence axiom is applied *locally* after every action is performed. Thus, if no exception to some properties is present while an action is performed the persistence axiom is used to conclude that those properties will remain unaltered in the resulting situation. Therefore, no property at the present is changed just for the sake of not changing some other properties in the future. The only reason for changes in properties are explicit changes in the properties by the action being performed at the moment.

It is straightforward to see that with this formalization of the persistence axiom DLEF precludes in a general and natural way unwanted extensions on temporal projection problems including the now (in)famous Yale Shooting Problem [H&M87].

Formalization of the Yale problem in DLEF

We shall adopt the notation of the situation calculus developed in [M&H69]. Consider the following axioms:

(1) $t(\text{ALIVE}, s_0)$

The person is alive in the initial situation.

(2) $\forall s.\ t(\text{LOADED}, result(\text{LOAD}, s)$

The gun becomes loaded any time a LOAD event happens.

(3) $\forall s.\ t(\text{LOADED}, s) \Rightarrow relevant(\text{ALIVE}, \text{SHOOT}, s)$

If the gun is loaded, then the SHOOT event is relevant to the fact that the person is alive.

(4) $\forall s.\ t(\text{LOADED}, s) \Rightarrow t(\neg\text{ALIVE}, result(\text{SHOOT}, s)$

If the gun is loaded, then the event SHOOT causes the person to be dead.

(5) $t(f, s) \Rightarrow t(f, result(e, s))$

(6) $relevant(f, e, s)\ //\ (t(f, s) \Rightarrow t(f, result(e, s)))$

(5) and (6) represent the axiom of persistence in DLEF. (5) asserts that a property *f* still holds after the preformance of action *e*, and (6) asserts that if the action *e* is relevant to property *f* (that is, *e* changes property *f*) than (5) is not applicable.

Let $T = \langle W, D, U \rangle$ be the following defeasible theory:

W = {(1), (2), (3), (4)}
D = {(5)}
U = {(6)}

Let S_1,S_2 and S_3 be the following situations:

S_1 = *result*(LOAD, S_0)
S_2 = *result*(WAIT, S_1)
S_3 = *result*(SHOOT, S_2)

The following axioms are closed instances of the axioms (1) to (6) above:

(1.1)	*t*(ALIVE, S_0)
(2.1)	*t*(LOADED, S_1)
(3.1)	*t*(LOADED, S_2) $\Rightarrow$ *relevant*(ALIVE, SHOOT, S_2)
(4.1)	*t*(LOADED, S_2) $\Rightarrow$ t($\neg$ALIVE, S_3)
(5.1)	*t*(ALIVE, S_1) $\Rightarrow$ *t*(ALIVE, S_2)
(5.2)	*t*(ALIVE, S_2) $\Rightarrow$ *t*(ALIVE, S_3)
(5.3)	*t*(LOADED, S_1) $\Rightarrow$ *t*(LOADED, S_2)
(5.4)	*t*(LOADED, S_2) $\Rightarrow$ *t*(LOADED, S_3)
(6.1)	*relevant*(ALIVE, SHOOT, S_2) // (*t*(ALIVE, S_2) $\Rightarrow$ *t*(ALIVE, S_3))

Observe that (5.3) $\angle$ (5.2), for (5.3) $\ll$ *relevant*(ALIVE, SHOOT, S_2), and *relevant*(ALIVE, SHOOT, S_2) // (5.2).

Therefore there is only one extension E = Th(W $\cup$ S) of *T* in DLEF, with {(5.1), (5.3), (5.4)} $\subseteq$ S. *E* corresponds to the expected descriptions for the situations S_0, ... ,S_3. In particular *t*(LOADED, S_2) $\in$ E and *t*($\neg$ALIVE, S_3) $\in$ E. Notice that E' = Th(W $\cup$ {(5.1), (5.2), (5.4)}) is not an extension for *T*.

6 Conclusions and Further Work

DLEF is a defeasible logic which prunes the set of extensions for a defeasible theory by ordering the defeasible formulae according to the order induced by the exceptions upon these formulae. The closest nonmonotonic logics to DLEF are Reiter's Default Logic

[Rei80], Delgrande's and Jackson's PJ-Logic [D&J91] and Poole's Theorist [Poo88]. We show in [Peq94] that these logics coincide for normal prerequisites-free theories — theories where exceptions are not explicitly represented — that is, E is an extension for a defeasible theory in DLEF iff E is an extension for the corresponding theories in these logics. However, in the presence of exceptions DLEF has less extensions than the other logics, we prove that any extension of DLEF — actually of a variant of DLEF called GDLEF for PJ-Logic and Theorist — is also an extension in these logics, but the inverse is not necessarily valid. A comprehensive comparison of DLEF with these logics is done in [Peq94].

A main application of DLEF is on the formalization of the frame problem. We showed in section 5 how DLEF implements the persistence axiom in a stepwise way.

Further developments to DLEF are in order. At the moment we investigate how the implicit hierarchy imposed by the order induced by the exceptions relates to the hierarchic nonmonotonic logics of Brewka [Bre89] and Konolige [Kon88].

References

[Bre89] Brewka, G., *Nonmonotonic Reasoning: From Theoretical Foundations to Efficient Computation.* PhD Thesis, 1989.

[D&J91] Delgrande, J. and Jackson, W., *Default Logic Revisited.* Proceedings of the Second international Conference on Principles of Knowledge Representation and Reasoning, Cambridge, Massachusetts, 1991, pp. 118-127.

[End72] Enderton, H.B., *A Mathematical Introduction to Logic,* Academic Press, New York, 1972.

[Eth88] Etherington, D.W., *Reasoning with Incomplete Information,* Morgan Kaufmann Publishers, London, 1988.

[H&M87] Hanks, S. & McDermott, D., *Nonmonotonic Logic and Temporal Projection,* Artificial Intelligence 33, 1987, pp.379-412.

[Kon88] Konolige, K., *Hierarchic Autoepistemic Theories for Nonmonotonic Reasoning.*Proceedings of the Seventh American Association for Artificial Intelligence Conference, St. Paul, Minnesota, 1988.

[M&H69] McCarthy, J. & Hayes, P.J., *Some philosophical problems from the standpoint of artificial intelligence,* D. Michie (Ed.), Machine Intelligence 4, American Elsevier, New York, 1969.

[Peq94] Pequeno, M., *Defeasible Reasoning with Exceptions First.* PhD Thesis, Department of Computing, Imperial College, University of London, 1994.

[Peq92] Pequeno, M.C., *Reasoning with Exceptions First.* In Proceedings of the XVIII Conferencia Latinoamericana de Informatica, CLEI 92, Las Palmas, Spain, 1992, pp 953-960.

[Peq91] Pequeno, M.C., *Three Principles for the Formalization of Nonmonotonic Reasoning,* in Proceedings of the Eighth Brazylian Symposium on Artificial Intelligence, Brasilia, November, 1991.

[Peq89] Pequeno, M.C., *Normally Canaries Fly*, In: Proceedingss of the Sixth Brazilan Symposium on Artificial Intelligence, Rio de Janeiro, November, 1989.

[Poo88] Poole, D.L, *A Logical Framework for Default Reasoning,* Artificial Intelligence **36**, 1988, pp. 27-47.

[Rei80] Reiter, R., *A logic for default reasoning,* Artificial Intelligence **13**, 1980, pp. 81-132.