# An assumption-based framework for non-monotonic reasoning [1]

**Andrei Bondarenko** [2]
Programming Systems Institute, Russian Academy of Sciences
Pereslavle-Zalessky, Russia
andrei@troyka.msk.su

**Francesca Toni, Robert A. Kowalski**
Department of Computing, Imperial College
180 Queen's Gate, London SW7 2BZ, UK
{ft, rak}@doc.ic.ac.uk

## Abstract

The notion of assumption-based framework generalises and refines the use of abduction to give a formalisation of non-monotonic reasoning. In this framework, a sentence is a non-monotonic consequence of a theory if it can be derived monotonically from a theory extended by means of acceptable assumptions. The notion of acceptability for such assumptions is formulated in terms of their ability successfully to "counterattack" any "attacking" set of assumptions. One set of assumptions is said to "attack" another if the first set monotonically implies a consequence which is inconsistent with an assumption in the second set. This argumentation-theoretic criterion of acceptability is based on notions first introduced for logic programming and used to give a unified account of such diverse semantics for logic programming as stable models, partial stable models, preferred extensions, stable theories, well-founded semantics, and stationary semantics. The new framework makes it possible to generalise various improvements first introduced for the semantics of logic programming and to apply these improvements to other formalisms for non-monotonic reasoning.

The paper investigates applications of the framework to logic programming, abductive logic programming, logic programs extended with "classical" negation, default logic, autoepistemic logic, and non-monotonic modal logic.

## 1 Introduction

In this paper we define a generalised framework for assumption-based reasoning and show how it can be applied both to logic programming and to

[1]Appeared in *Proc. 2nd International Workshop on Logic Programming and Non-monotonic Reasoning*, (A. Nerode and L. Pereira eds.) MIT Press (1993).

[2]This work was done while the author was a visitor at Imperial College

other formalisms for non-monotonic reasoning. The new framework was inspired by Dung's general argumentation framework [6], but is formulated differently as a generalisation of the abductive frameworks of Poole [16] and Eshghi and Kowalski [7].

The new framework generalises the approach of [16] and shows how any monotonic logic can be extended to a non-monotonic logic by appropriately identifying a set of candidate assumptions and specifying the conditions under which a theory can be extended by an acceptable set of assumptions. It replaces the notion that a set of assumptions is acceptable if it is consistent with the theory by the more refined notion that it is acceptable if it is consistent with the theory and can "counterattack" any "attacking" set of assumptions. A set of assumptions is said to "attack" another if together with the theory it implies a consequence which is inconsistent with some assumption contained in the other set. Like Dung's argumentation-theoretic framework, the assumption-based framework investigated in this paper generalises the notion of attacking and counterattacking sets of assumptions introduced for logic programming by Kakas, Kowalski and Toni [13]. It also defines a new notion of counterattack which improves upon previous definitions.

The framework makes it possible to generalise various improvements first introduced for the semantics of logic programming and to apply these improvements to other formalisms for non-monotonic reasoning. In this paper we will propose such improvements specifically for a number of existing approaches to abductive logic programming, extended logic programming, default logic, autoepistemic logic and non-monotonic modal logic.

The paper is organised into four main parts: The first introduces the general assumption-based framework and the improved notion of counterattack, the second applies the framework to logic programming, the third to abductive logic programming and to logic programming extended with so-called "classical" negation, and the fourth part to default logic, autoepistemic logic and non-monotonic modal logic.

## 2 Basic definitions

In this paper, a *deductive system* is a pair $(\mathcal{L}, \mathcal{R})$ such that

- $\mathcal{L}$ is a formal language with a special formula $\bot \in \mathcal{L}$, denoting falsity,
- $\mathcal{R}$ is a set of inference rules of the form

$$\frac{\alpha_1, \ldots, \alpha_n}{\alpha}$$

  where $\alpha, \alpha_1, \ldots, \alpha_n \in \mathcal{L}$ and $n \geq 0$.

Notice that logical axioms can be represented as inference rules with $n = 0$. Any set of formulae $T \subseteq \mathcal{L}$ is called a *theory*.

A *deduction* from a theory $T$ is a sequence $\beta_1, \ldots, \beta_m$, where $m > 0$, such that, for all $i = 1, \ldots, m$,

- $\beta_i \in T$, or
- there exists $\dfrac{\alpha_1, \ldots, \alpha_n}{\beta_i}$ in $\mathcal{R}$ such that $\alpha_1, \ldots, \alpha_n \in \{\beta_1, \ldots, \beta_{i-1}\}$.

$T \vdash \alpha$ means that there is a deduction from $T$ whose last element is $\alpha$. A theory $T$ is said to be *inconsistent* if $T \vdash \perp$, and *consistent* otherwise.
$Th(T)$ is the set $\{\alpha \in \mathcal{L} \,|\, T \vdash \alpha\}$.

**Definition 2.1** An *assumption-based framework*
is a pair $\langle(\mathcal{L}, \mathcal{R}), Ab\rangle$ such that

- $(\mathcal{L}, \mathcal{R})$ is a deductive system,
- $Ab \subseteq \mathcal{L}$.

The elements of $Ab$ are called *assumptions* and the set $Ab$ represents the set of all candidate assumptions that can be used to extend a given theory. Notice that deductive systems are monotonic. Non-monotonicity arises because a set of assumptions which acceptably extends a given theory may be unacceptable for a larger theory.

The notion of assumption-based framework can be viewed as a direct generalisation of Poole's abductive framework. Whereas he considers only the deductive system of first-order logic, we admit deductive systems for any monotonic logic. Moreover, whereas Poole allows a set of assumptions to extend a theory if it is consistent with the theory, we allow such an extension if it is consistent with the theory and can successfully counterattack any attack.

In this section we will assume that an arbitrary but fixed assumption-based framework $\langle(\mathcal{L}, \mathcal{R}), Ab\rangle$ is given.

**Definition 2.2** Given a theory $T$ and sets of assumptions $\Delta$ and $\mathcal{A}$,
$\mathcal{A}$ *attacks* $\Delta$ (with respect to $T$) if and only if
there exist $\alpha \neq \perp$ and $\beta \in \Delta$ such that
$T \cup \mathcal{A} \vdash \alpha$, and
$\{\alpha, \beta\} \vdash \perp$.

In other words, $\mathcal{A}$ *attacks* $\Delta$ with respect to a theory $T$ if there is a deduction from $T \cup \mathcal{A}$ which contradicts one of the assumptions in $\Delta$. This deduction can be regarded as an *argument* against $\Delta$, based upon the assumptions in $\mathcal{A}$. In the sequel, we will normally omit the qualification "with respect to $T$" when the identity of $T$ is clear from the context.

A consistent set of assumptions $\Delta$ is admissible (or acceptable) if it can counterattack any set of assumptions $\mathcal{A}$ that *attacks* it. Before we define admissibility more formally, we need to define the notion of *counterattack*. Several alternative notions of *counterattack* will be investigated in this paper. The following definition presents the most important of these.

**Definition 2.3** Given a theory $T$ and sets of assumptions $\Delta$ and $\mathcal{A}$,

1) $\Delta$ *counterattacks*$_1$ $\mathcal{A}$ if and only if
   $\Delta$ *attacks* $\mathcal{A}$;

2) $\Delta$ *counterattacks*$_2$ $\mathcal{A}$ if and only if
   $\Delta$ *attacks* $\mathcal{A}$ or
   $T \cup \mathcal{A}$ is inconsistent.

The definitions below are all relative to the notion of *counterattack*, and are adapted from those given by Dung [6]. We will argue in this paper that the first two definitions, of admissible and preferred sets of assumptions, can provide the basis for an improved semantics for non-monotonic reasoning in general.

**Definition 2.4** A set of assumptions $\Delta$ is *admissible*
(with respect to a theory $T$) if and only if

- $T \cup \Delta$ is consistent, and
- for all sets of assumptions $\mathcal{A}$,
  if $\mathcal{A}$ *attacks* $\Delta$, then $\Delta$ *counterattacks* $\mathcal{A}$.

Note that the empty set of assumptions is admissible with respect to any consistent theory.

**Definition 2.5** A set of assumptions $\Delta$ is *preferred*
(with respect to a theory $T$) if and only if
$\Delta$ is maximally (with respect to set inclusion) admissible.

It is easy to see that, for every consistent theory, there always exists a set of assumptions which is preferred.

The following two definitions, of complete and grounded sets of assumptions, provide the basis for a sceptical semantics. Informally, a consistent set of assumptions is complete if it consists of all the assumptions that it defends, where it defends an assumption if it *counterattacks* any attack against that assumption. A set of assumptions is grounded if it is minimally complete. In logic programming the notion of groundedness corresponds to the well-founded semantics [20, 4].

**Definition 2.6** A set of assumptions $\Delta$ is *complete*
(with respect to a theory $T$) if and only if

- $T \cup \Delta$ is consistent, and
- $\Delta = \{\alpha \mid \alpha \in Ab$ and
  $\forall\ \mathcal{A} \subseteq Ab$, if $\mathcal{A}$ *attacks* $\{\alpha\}$, then $\Delta$ *counterattacks* $\mathcal{A}\ \}$.

**Definition 2.7** A set of assumption $\Delta$ is *grounded*
(with respect to a theory $T$) if and only if
$\Delta$ is minimally (with respect to set inclusion) complete.

The following definition, of stable set of assumptions, provides the basis for a credulous semantics. As we will see later in this paper, this semantics corresponds to many of the semantics which have been proposed for different formalisms for non-monotonic reasoning, including the stable model semantics of logic programming [8], and extensions in default logic [18], autoepistemic logic [15] and non-monotonic modal logic [14]. Intuitively, a consistent set of assumptions is stable if it *attacks* every assumption it does not contain.

**Definition 2.8** A set of assumptions $\Delta$ is *stable*
(with respect to a theory $T$) if and only if

- $T \cup \Delta$ is consistent, and
- $\forall \alpha \in Ab$, if $\alpha \notin \Delta$ then $\Delta$ *attacks* $\{ \alpha \}$.

The following three properties are direct consequences of the definitions and do not depend upon the *counterattacks* relation. Given a theory $T$ and a set of assumptions $\Delta$:

- If $\Delta$ is preferred then $\Delta$ is admissible.
- If $\Delta$ is complete then $\Delta$ is admissible.
- If $\Delta$ is grounded then $\Delta$ is complete.

The following property holds for all the notions of *counterattack* defined in this paper:

- If $\Delta$ is stable then $\Delta$ is preferred.

To show how the notions defined in this section can be used to provide a uniform formulation of many existing approaches to non-monotonic reasoning, we will need the notion of extension given in the following definition.

**Definition 2.9** $E$ is a *preferred (stable, complete or grounded) extension*
of a consistent theory $T$ if and only if there exists (with respect to $T$)
a preferred (stable, complete or grounded, respectively)
set of assumptions $\Delta$ such that $E = Th(T \cup \Delta)$.
$E$ is a *preferred (stable, complete or grounded) extension*
of an inconsistent theory $T$ if and only if $E = Th(T)$.

We will argue that admissible and preferred sets of assumptions with $counterattacks_2$ provide a better semantics for non-monotonic reasoning than either stable sets or admissible sets with $counterattacks_1$. Consequently, we introduce the notions of *weakly admissible* and *weakly preferred sets* and *weakly preferred extensions* to make it easier to refer to these notions later in the paper:

**Definition 2.10** A set of assumptions $\Delta$ is *weakly admissible* (with respect to a theory $T$) if and only if $\Delta$ is admissible with $counterattacks_2$.

**Definition 2.11** A set of assumptions $\Delta$ is *weakly preferred* (with respect to a theory $T$) if and only if $\Delta$ is preferred with $counterattacks_2$.

**Definition 2.12** $E$ is a *weakly preferred extension* of a theory $T$ if and only if $E$ is a preferred extension of $T$ with $counterattacks_2$.

As mentioned in the introduction of this paper, our notion of assumption-based framework was inspired by Dung's argumentation-based framework [6]. The role of assumptions in our approach is played by (abstract) arguments in Dung's approach. On the one hand, assumptions can be viewed mathematically as a special case of arguments; on the other hand, arguments can be understood in our framework as deductions from a theory extended with assumptions.

Dung's notion of *attack* is more abstract than ours. We have attempted to identify notions of *attack* and *counterattack* which are as specific as possible, but also general enough to capture as many existing approaches to non-monotonic reasoning as possible. Later, when we investigate autoepistemic and non-monotonic modal logics, we will extend our framework to include a notion of preference to capture better the semantics of these logics.

# 3 Logic programming

We will assume that the semantics of a logic program containing variables is given by the set of all its variable-free instances over some Herbrand universe. $\mathcal{HB}$ will stand for the *Herbrand base* of variable-free atoms formulated over this Herbrand universe, $\mathcal{HB}_{not}$ will stand for the set $\{not\, p \mid p \in \mathcal{HB}\}$ and $Lit$ will stand for $\mathcal{HB} \cup \mathcal{HB}_{not}$.

The assumption-based framework for logic programming is $\langle(\mathcal{L}, \mathcal{R}), Ab\rangle$ where

- $\mathcal{L} = \{\bot\} \cup Lit \cup \{p \leftarrow l_1, \ldots, l_n \mid p \in \mathcal{HB}, l_1, \ldots, l_n \in Lit, \text{ and } n \geq 0\}$,
- $\mathcal{R}$ is the set of all inference rules of the form

$$\frac{p \leftarrow l_1, \ldots, l_n, \qquad l_1, \ldots, l_n}{p}$$

  where $p \in \mathcal{HB}$, $l_1, \ldots, l_n \in Lit$, and $n \geq 0$, and of the form

$$\frac{p, \qquad not\, p}{\bot}$$

  where $p \in \mathcal{HB}$,

- $Ab = \mathcal{HB}_{not}$.

A logic program $P$ is a theory, $P \subseteq \mathcal{L}$, in such an assumption-based framework.

The interpretation of negative literals as abducibles was first presented in [7], and was the basis for the preferred extension semantics [4], the stable theory and acceptability semantics [11], and the argumentation-theoretic interpretation for the semantics of logic programming presented in [13].

The instance of the definition 2.2 of *attack* for the assumption-based framework $\langle(\mathcal{L}, \mathcal{R}), Ab\rangle$ for logic programming is the following:

- Given a logic program $P$ and sets of assumptions $\Delta$ and $\mathcal{A}$, $\mathcal{A}$ *attacks* $\Delta$ if and only if $P \cup \mathcal{A} \vdash p$, for some $not\, p \in \Delta$.

Note that this definition coincides with that presented in [13].

By instantiating the different definitions presented in section 2 with respect to the assumption-based framework for logic programming with $counterattacks_1$, we can obtain different existing semantics for negation as failure.

**Theorem 3.1** *Given a logic program P, and* $counterattacks_1$ *as the definition of counterattacks,*

(a) *M is a stable model [8] of P*
*if and only if*
*there is a stable extension E of P, such that* $M = E \cap \mathcal{HB}$*;*

(b) *given a set of assumptions* $\Delta$*,*
$P \cup \Delta$ *is a complete scenario [4]*
*(and* $Th(P \cup \Delta) \cap Lit$ *is a well-founded model [20]) of P*
*if and only if*
$P \cup \Delta$ *is a stationary expansion [17] of P*
*if and only if*
$\Delta$ *is complete (grounded respectively) with respect to P;*

(c) *given a set of assumptions* $\Delta$*,*
$P \cup \Delta$ *is a preferred extension in the sense of [4]*
*(and* $P \cup \Delta$ *is an admissible scenario [4]) of P*
*if and only if*
$Th(P \cup \Delta) \cap\ Lit$ *is a partial stable model [19] of P*
*if and only if*
$\Delta$ *is preferred (*$\Delta$ *is admissible respectively) with respect to P.*

This theorem is an immediate consequence of results presented by Dung in [4, 6], together with results in [3] and [12].

The following example shows that preferred extensions are better than stable models.

**Example 3.1** The program

$$\{p \leftarrow not\, p\}$$

has no stable extension, but it has a preferred extension corresponding to the preferred set of assumptions $\emptyset$. Preferred extension semantics is consequently more modular than stable model semantics. For example, the program

$$\{q, \quad p \leftarrow not\, p\}$$

has no stable extension, but it has a preferred extension containing $q$.

### 3.1 Stable theories and acceptability semantics

To capture stable theory and acceptability semantics [11] we need two new notions of *counterattack*, different from those introduced in definition 2.3. For simplicity, we present these notions in the assumption-based framework for logic programming. However, they can also be defined more generally and can be applied to any other assumption-based framework.

**Definition 3.1** Given a logic program $P$ and sets of assumptions $\Delta$ and $\mathcal{A}$,
$\Delta$ *counterattacks*$_3$ $\mathcal{A}$ if and only if
$\Delta \cup \mathcal{A}$ *attacks* $\mathcal{A} \perp \Delta$ [3].

The following theorem is an immediate consequence of definition 3.1 and the definitions given in [11].

**Theorem 3.2** *Given a program $P$,*
*a set of assumptions $\Delta$ is weakly stable [11]*
*(and $P \cup \Delta$ is a stable theory [11]) with respect to $P$*
*if and only if*
*$\Delta$ is admissible (preferred respectively)*
*with respect to $P$ with counterattacks$_3$.*

The following example shows that *counterattacks*$_3$ is "better" than *counterattacks*$_1$.

**Example 3.2** The program

$$\{q \leftarrow not\, p, \quad p \leftarrow not\, p\}$$

has only one admissible set of assumptions, $\emptyset$, with *counterattacks*$_1$. However, it has the admissible set $\{not\, q\}$ with *counterattacks*$_3$, because the only attack against it, $\{not\, p\}$, is inconsistent.

[3] In the published version of this paper *counterattacks*$_3$ was defined by
$\Delta \cup \mathcal{A}$ *attacks* $\mathcal{A}$.
As noted by Noboru Iwayama, with this definition theorem 3.2 does not hold.

The acceptability semantics was introduced in [11] to overcome certain disadvantages of stable theories. Before presenting the definition, we note that the notion of admissibility could have been defined more generally, relative to an already accepted set of assumptions.

**Definition 3.2** Given a logic program $P$, sets of assumptions $\Delta$ and $\Delta_0$, and a specific definition of the *counterattacks* relation,
$\Delta$ is *acceptable to* $\Delta_0$ if and only if
for all sets of assumptions $\mathcal{A}$,
if $\mathcal{A}$ *attacks* $\Delta \perp \Delta_0$, then $\Delta \cup \Delta_0$ *counterattacks* $\mathcal{A}$.

**Definition 3.3** Given a logic program $P$ and sets of assumptions $\Delta$ and $\mathcal{A}$,
$\Delta$ *counterattacks*$_4$ $\mathcal{A}$ if and only if
$\mathcal{A}$ is not acceptable to $\Delta$ with *counterattacks*$_4$.

Notice that definition 3.3 is recursive and that definition 3.2 becomes recursive with *counterattacks*$_4$.

The following theorem is an immediate consequence of definitions 3.2 and 3.3 and the definition of acceptability given in [11].

**Theorem 3.3** *Given a program $P$ and sets of assumptions $\Delta$ and $\Delta_0$, $\Delta$ is acceptable to $\Delta_0$ according to [11] with respect to $P$ if and only if $\Delta$ is acceptable to $\Delta_0$ with respect to $P$ with counterattacks$_4$.*

Note that, given a logic program $P$ and a set of assumptions $\Delta$, if $\Delta$ is admissible in the sense of [4] then $\Delta$ is weakly stable [11], and if $\Delta$ is weakly stable then $\Delta$ is acceptable to $\emptyset$ in the sense of [11] (see [13]).

## 3.2 Improved semantics

In this section we will illustrate the new semantics for negation as failure in logic programming, based on the notions of weakly admissible and weakly preferred sets of assumptions introduced in definitions 2.10 and 2.11. This new semantics can be understood as an improvement of the stable theory semantics, as demonstrated by the following example.

**Example 3.3** The logic program

$$P = \{p \leftarrow not\, q, \quad q \leftarrow not\, p, not\, q\}$$

has two preferred sets of assumptions with *counterattacks*$_3$, $\{not\, q\}$ and $\{not\, p\}$. The second set $\Delta = \{not\, p\}$ can *counterattack*$_3$ the attack $\mathcal{A}=\{not\, q\}$, because $q$ can be derived from the combined attack $\Delta \cup \mathcal{A}=\{not\, p,\, not\, q\}$. But it can be argued that this combined attack should not be accepted because it is inconsistent with $P$. Its assumptions are held neither by the defendant $\Delta$ nor by the prosecutor $\mathcal{A}$. The notion of weakly preferred set of assumptions (where *counterattacks*$_2$ replaces *counterattacks*$_3$) gives the intuitively correct result, only $\{not\, q\}$, in this example.

Note that, given a logic program $P$, if a set of assumptions $\Delta$ is admissible in the sense of [4] then $\Delta$ is weakly admissible and weakly stable [11] [4].

In the same way we improve $counterattacks_3$ by $counterattacks_2$, we can improve $counterattacks_4$ by a new notion of acceptability. This is, however, beyond the scope of this paper.

# 4 Extensions of logic programming

## 4.1 Abductive logic programming

An abductive logic program is a triple $\langle P, Ab_0, I\rangle$, where $P$ is a logic program, $Ab_0$ is a set of variable-free atoms representing a set of abducibles, and $I$ is a set of closed first-order formulas, representing integrity constraints. Without lost of generality (see [13]) we assume that integrity constraints are represented as clauses of the form

$$\bot \leftarrow l_1, \ldots, l_n.$$

We also assume that $Ab_0$ is disjoint from the conclusions of clauses in $P$.

As in section 3, we will assume that logic programs and integrity constraints containing variables represent all their variable-free instances over some Herbrand universe. Consequently, we will assume that programs and constraints are variable-free.

The assumption-based framework corresponding to a set of abducibles $Ab_0$ is $\langle(\mathcal{L}, \mathcal{R}), Ab\rangle$ where

- $\mathcal{L}$ is the language of section 3, extended by all clauses of the form

$$\bot \leftarrow l_1, \ldots, l_n$$

  where $l_1, \ldots, l_n \in Lit$ and $n \geq 0$,

- $\mathcal{R}$ is the set of inference rules of section 3,

- $Ab = Ab_0 \cup \mathcal{HB}_{not}$.

The notion of *attacks* presented in definition 2.2 can be written as:

- Given an abductive logic program $\langle P, Ab_0, I\rangle$
  and sets of assumptions $\Delta$ and $\mathcal{A}$
  in the assumption-based framework corresponding to $Ab_0$,
  $\mathcal{A}$ *attacks* $\Delta$ (with respect to $P \cup I$) if and only if
  $P \cup \mathcal{A} \vdash p$, for some $not\, p \in \Delta$, or
  $not\, a \in \mathcal{A}$, for some $a \in \Delta \cap Ab_0$.

[4]In the published version of this paper we also claimed that if $\Delta$ is weakly admissible then $\Delta$ is weakly stable [11], which is a mistake.

Note that in this approach the integrity constraints $I$ are used only to check consistency, and not to create attacks.

The generalised stable model semantics of [10] is a special case of the general definition of stability.

**Theorem 4.1** *Given an abductive logic program* $\langle P, Ab_0, I\rangle$, *$M$ is a generalised stable model [10] of* $\langle P, Ab_0, I\rangle$ *if and only if there is a stable extension $E$ of $P \cup I$ in the assumption-based framework corresponding to $Ab_0$ and* $M = E \cap \mathcal{HB}$.

## 4.2 Improved semantics

The generalised stable model semantics inherits the disadvantages of the stable model semantics illustrated in example 3.1. As in the case of normal logic programs, many of these disadvantages can be overcome by replacing stable models by preferred extensions. However, this one change alone (leaving $counterattacks_1$ unchanged) does not overcome all problems, as shown by the following example.

**Example 4.1** Let the abductive program $\langle P, Ab_0, I\rangle$ be given by

$$\begin{array}{lcl} P & = & \{p \leftarrow a\} \\ Ab_0 & = & \{a\} \\ I & = & \{\bot \leftarrow a, \quad \bot \leftarrow not\, a\}. \end{array}$$

Intuitively, the set of assumptions $\Delta = \{not\, p\}$ should be admissible with respect to $P \cup I$, because $p$ cannot hold. But $\Delta$ is not admissible with $counterattacks_1$ because $\mathcal{A}=\{a\}$ *attacks* $\Delta$, but $\Delta$ does not $counterattack_1$ $\mathcal{A}$. However, $\Delta$ does $counterattack_2$ $\mathcal{A}$, because $\mathcal{A}$ is inconsistent.

The notions of weakly admissible and weakly preferred sets of assumptions give the intuitively correct result in this and similar examples.

## 4.3 Extended logic programming

Extended logic programming is the extension of logic programming to incorporate explicit negation in addition to negation as failure. As in the case of normal logic programs and abductive logic programs, we will assume that extended logic programs are variable-free. $\mathcal{HB}$ will stand for the Herbrand base, $\mathcal{HB}_e$ will stand for $\mathcal{HB} \cup \{\neg\, p \,|\, p \in \mathcal{HB}\ \}$ and $Lit_e$ will stand for $\mathcal{HB}_e \cup \{not\, l \,|\, l \in \mathcal{HB}_e\ \}$.

The assumption-based framework for extended logic programming is $\langle(\mathcal{L}, \mathcal{R}), Ab\rangle$ where

- $\mathcal{L} = \{\bot\} \cup Lit_e \cup$
  $\{l \leftarrow l_1, \ldots, l_n \mid l \in \mathcal{HB}_e,\ l_1, \ldots, l_n \in Lit_e, \text{ and } n \geq 0\}$,

- $\mathcal{R}$ is the set of all inference rules of the form

$$\frac{l \leftarrow l_1, \ldots, l_n, \quad l_1, \ldots, l_n}{l}$$

where $l \in \mathcal{HB}_e$, $\quad l_1, \ldots, l_n \in Lit_e$, and $n \geq 0$, and of the form

$$\frac{l, \quad\quad not\, l}{\bot}$$

where $l \in \mathcal{HB}_e$,

- $Ab = \{not\, l \mid l \in \mathcal{HB}_e\}$.

The negation denoted by $\neg$ is called "classical" negation in [9]. However, in this paper we use the term "explicit" negation, because clauses of extended logic programs are treated more like inference rules than like classical implications.

The instance of the definition 2.2 of *attack* for extended logic programming is identical to the definition for logic programming, except that $\mathcal{HB}$ is replaced by $\mathcal{HB}_e$.

#### 4.3.1 Answer set semantics

The answer set semantics [9] is a special case of the general definition of stability where extended logic programs are extended by the further clauses $\{l \leftarrow p, \neg p \,|\, l \in \mathcal{HB}_e$, and $p \in \mathcal{HB}\}$. As a result, this semantics is classical only in the sense that from an inconsistency any conclusion can be derived.

**Theorem 4.2** *Given an extended logic program $P$,*
*$M$ is an answer set [9] of $P$ if and only if*
*there is a stable extension $E$ of the theory*
*$P \cup \{l \leftarrow p, \neg p \,|\, l \in \mathcal{HB}_e$, and $p \in \mathcal{HB}\}$ in the corresponding*
*assumption-based framework and $M = E \cap \mathcal{HB}_e$.*

#### 4.3.2 The Dung and Ruamviboonsuk semantics

Dung and Ruamviboonsuk's semantics [5] is a special case of admissibility semantics with $counterattacks_1$ where extended logic programs are further extended by the integrity constraints $\{\bot \leftarrow p, \neg p \,|\, p \in \mathcal{HB}\}$.

**Theorem 4.3** *Given an extended logic program $P$ and a set of assumptions*
*$\Delta$, $P \cup \Delta$ is an admissible scenario [5] of $P$ if and only if*
*$\Delta$ is admissible with respect to $P \cup \{\bot \leftarrow p, \neg p \,|\, p \in \mathcal{HB}\}$*
*in the corresponding assumption-based framework with $counterattacks_1$.*

### 4.3.3 Improved semantics

As in other cases, the admissibility (and preferred extension) semantics with $counterattacks_1$ sometimes gives intuitively incorrect results, as illustrated by the following example.

**Example 4.2** Consider the extended logic program

$$\{\neg p, \quad p \leftarrow not\, q\}$$

further extended by the integrity constraints

$$\{\bot \leftarrow p, \neg p \mid p \in \mathcal{HB}\}.$$

The set of assumptions $\{not\, p\}$ is not admissible with $counterattacks_1$ because $\{not\, q\}$ *attacks* $\{not\, p\}$ but cannot be $counterattacked_1$ by $\{not\, p\}$. Intuitively, however, the theory should have a preferred extension in which $not\, p$ holds, because the attack $\{not\, q\}$ is inconsistent with the theory. This extension can be obtained by using weakly preferred extensions (with $counterattacks_2$ instead of $counterattacks_1$).

The definitions 2.10 and 2.11 in this case become:

- Given an extended logic program $P$,
  a set of assumptions $\Delta$ is *weakly admissible* (*weakly preferred*)
  if and only if $\Delta$ is admissible (preferred respectively)
  with respect to $P \cup \{\bot \leftarrow p, \neg p \mid p \in \mathcal{HB}\}$ in the corresponding
  assumption-based framework with $counterattacks_2$.

# 5 Default logic

Let $(\mathcal{L}_0, \mathcal{R}_0)$ be a deductive system for first-order logic, where $\mathcal{L}_0$ contains a special formula, $\bot$, denoting falsity. Following [18], a default theory is a pair $(T, D)$ where

- $T \subseteq \mathcal{L}_0$,
- $D$ is a set of default rules of the form

$$\frac{\alpha : M\beta_1, \ldots, M\beta_n}{\gamma}$$

  where $\alpha$, $\beta_1, \ldots, \beta_n$, $\gamma \in \mathcal{L}_0$, and $n \geq 0$.

We will assume that all defaults rules in $D$ are closed, i.e. they contain no free variables. (As in logic programming, default rules containing free variables represent all their variable-free instances.)

The assumption-based framework corresponding to $D$ in such a default theory $(T, D)$ is $\langle(\mathcal{L}, \mathcal{R}), Ab\rangle$ where

- $\mathcal{L} = \mathcal{L}_0 \cup \{M\,\phi \,|\, \phi \in \mathcal{L}_0$ and $\phi$ is closed$\}$,
- $\mathcal{R}$ is $\mathcal{R}_0$ extended with the set of all inference rules of the form

$$\frac{\alpha,\ M\beta_1, \ldots, M\beta_n}{\gamma}$$

  where

$$\frac{\alpha : M\beta_1, \ldots, M\beta_n}{\gamma} \in D,$$

  and of the form

$$\frac{\neg\phi,\ M\,\phi}{\bot}$$

  where $\phi \in \mathcal{L}_0$,
- $Ab = \{M\,\phi \,|\, \phi \in \mathcal{L}_0$ and $\phi$ is closed$\}$.

An assumption of the form $M\phi$ intuitively means that $\phi$ is consistent, i.e. that $\neg\phi$ can not be derived.

The notion of *attack* presented in definition 2.2 becomes:

- Given a default theory $(T,\ D)$ and sets of assumptions $\Delta$ and $\mathcal{A}$, in the assumption-based framework corresponding to $D$, $\mathcal{A}$ *attacks* $\Delta$ if and only if $T \cup \mathcal{A} \vdash \neg\phi$, for some $M\phi \in \Delta$.

The following theorem is a consequence of a theorem in [1].

**Theorem 5.1** *$E$ is an extension [18] of a default theory $(T,\ D)$ if and only if there is a stable extension $E'$ of $T$ in the assumption-based framework corresponding to $D$ and $E = E' \cap \mathcal{L}_0$.*

As we have already seen earlier in this paper, the notion of stability is sometimes too strong. This is illustrated for default logic by the following example, which is a variant of example 3.1.

**Example 5.1** The default theory $(T,\ D)$ where $T = \emptyset$ and

$$D = \left\{ \frac{: M\neg p}{p},\ \frac{: Mq}{q} \right\}$$

has no extension in Reiter's default logic. However, intuitively it should have an extension containing $q$.

The problem with this example can be solved by replacing stability by admissibility, without changing $counterattacks_1$. However, $counterattacks_1$ gives other problems in other examples, because of the fact that in first-order logic an inconsistent set of assumptions implies every sentence. Therefore, in default logic an inconsistent set of assumptions *attacks* every non-empty set of assumptions. This is illustrated by the following example.

**Example 5.2** The default theory $(T, D)$ where $T = \{\neg p\}$ and

$$D = \left\{ \frac{: Mr}{p}, \frac{: Mq}{q} \right\}$$

should intuitively have an extension containing $q$. However, the only preferred set of assumptions is $\emptyset$ in the corresponding assumption-based framework with *counterattacks*$_1$. This is because $\{Mr\}$ is inconsistent with $T$ and therefore implies $\neg q$ and *attacks* $\{Mq\}$. But $\{Mq\}$ does not *counterattack*$_1$ $\{Mr\}$. Replacing *counterattacks*$_1$ by *counterattacks*$_2$ (and therefore preferred extensions by weakly preferred extensions) we obtain the intuitively correct result.

In the case of default logic, the definition 2.12 becomes:

- Given a default theory $(T, D)$,
  $E$ is a *weakly preferred* extension of $(T, D)$ if and only if
  $E$ is a preferred extension of $T$ in the assumption-based framework corresponding to $D$ with *counterattacks*$_2$.

# 6 Assumption-based framework with preferences

In this section we will present a generalisation of the assumption-based framework which includes the notion of preferences between formulae in the language. This will allow us to capture and to propose improvements for autoepistemic logic [15] and non-monotonic modal logic [14].

**Definition 6.1** An *assumption-based framework (with preferences)* is a triple $\langle(\mathcal{L}, \mathcal{R}), Ab, \leq\rangle$ such that

- $(\mathcal{L}, \mathcal{R})$ is a deductive system (with $\perp \in \mathcal{L}$),
- $Ab \subseteq \mathcal{L}$,
- $\leq \subseteq \mathcal{L} \times \mathcal{L}$.

Intuitively, $p \leq q$ means that if $p$ and $q$ can not hold together then $p$ should be preferred to $q$.

**Definition 6.2** Given a theory $T$ and sets of assumptions $\Delta$ and $\mathcal{A}$,
$\mathcal{A}$ *attacks* $\Delta$ (with respect to $T$) if and only if
there exist $\alpha \neq \perp$ and $\beta \in \Delta$ such that
$T \cup \mathcal{A} \vdash \alpha$,
$\{\alpha, \beta\} \vdash \perp$ and
$\alpha \leq \beta$.

Given this new definition of *attack*, the definitions of *counterattack* given in section 3 remain unchanged, as do the definitions of *admissible, preferred, complete, grounded* and *stable sets* of assumptions and *extensions*.

Note that if $\leq\, = \mathcal{L} \times \mathcal{L}$, the condition $\alpha \leq \beta$ plays no role in the definition of *attack*, and therefore can be omitted. Consequently, the framework defined in section 3 is a special case of the framework presented here.

This framework is related to the extension of Poole's abductive framework introduced by Brewka [2]. One major difference between our approaches is that Brewka defines preference between abducibles whereas we define preferences more generally between formulae of the language. Further work is necessary to determine whether our framework can capture Brewka's approach or whether some further generalisation of our framework is necessary for this purpose.

### 6.1 Autoepistemic logic

Autoepistemic logic [15] is based upon a deductive system $(\mathcal{L}, \mathcal{R})$, where $\mathcal{L}$ is a propositional modal language containing a modality $L$, and $\mathcal{R}$ is some presentation of classical propositional logic for the language $\mathcal{L}$. The intended meaning of $L\phi$ is that $\phi$ is believed. As before, we assume that $\bot \in \mathcal{L}$.

Following [15], $E \subseteq \mathcal{L}$ is an *autoepistemic extension* of a *theory* $T \subseteq \mathcal{L}$ if and only if $E = Th(T \cup \{L\phi \mid \phi \in E\} \cup \{\neg L\phi \mid \phi \in \mathcal{L} \perp E\})$.

Autoepistemic logic can be formulated in terms of the assumption-based framework $\langle(\mathcal{L}, \mathcal{R}), Ab, \leq\rangle$ where

- $Ab = \{\neg L\phi \mid \phi \in \mathcal{L}\} \cup \{L\phi \mid \phi \in \mathcal{L}\}$,
- $\leq$ is defined by: $\phi \leq \neg L\phi$ and $\neg L\phi \leq L\phi$, for all $\phi \in \mathcal{L}$.

In this framework both positive and negative beliefs can be assumptions. Intuitively, the preference relation expresses that we always prefer to know whether or not a proposition $\phi$ holds, but if there is no such knowledge about $\phi$, and we have to make a choice between believing and not believing $\phi$, then we prefer to be sceptical, choosing $\neg L\phi$ over $L\phi$.

In the assumption-based framework corresponding to autoepistemic logic, the notion of *attack* becomes:

- Given a theory $T$ and sets of assumptions $\Delta$ and $\mathcal{A}$,
  $\mathcal{A}$ *attacks* $\Delta$ if and only if
  $T \cup \mathcal{A} \vdash \phi$, for some $\neg L\phi \in \Delta$, or
  $T \cup \mathcal{A} \vdash \neg L\phi$, for some $L\phi \in \Delta$.

Notice that $\{\neg L\phi\}$ *attacks* $\{L\phi\}$ but not vice versa.

**Theorem 6.1** *$E$ is an autoepistemic extension of a theory $T$ if and only if $E$ is a stable extension of $T$ in the corresponding assumption-based framework.*

As in the other cases investigated in this paper, stable extensions have a number of disadvantages compared with weakly preferred extensions.

**Example 6.1** The autoepistemic theory $\{\neg L\phi \supset \phi\}$, for example, similar to the logic program of example 3.1 and the default theory of example 5.1, has no stable extension but has an unique weakly preferred extension based upon the empty set of assumptions.

Similarly, the theory $\{L\phi\}$ has no stable extension but has a unique weakly preferred extension.

However, using weakly preferred extensions instead of stable extensions does not solve all the problems, as illustrated by the following example.

**Example 6.2** The theory $\{L\phi \supset \phi\}$ has two stable and weakly preferred extensions, one containing the assumption $\neg L\phi$, the other containing $L\phi$. The second extension is anomalous.

One way to avoid the anomalous extension is to restrict assumptions to negative beliefs, $\neg L\phi$, and to express positive introspection by means of a new inference rule

$$\frac{\phi}{L\phi}.$$

For this purpose, we need to replace the deductive system $(\mathcal{L}, \mathcal{R})$ for autoepistemic logic by one where $\mathcal{R}$ is based upon modal rather than classical logic, as in non-monotonic modal logic.

### 6.2 Non-monotonic modal logic

Non-monotonic modal logic [14] can be formulated in terms of a deductive system $(\mathcal{L}, \mathcal{R})$ where $\mathcal{L}$ is a first-order modal language containing a modal operator, $L$, and a special formula, $\perp$, and where $\mathcal{R}$ is some presentation of a modal system for the language $\mathcal{L}$, containing all instances of the *necessitation rule* of inference:

$$\frac{\phi}{L\phi} \qquad \text{for all } \phi \in \mathcal{L}.$$

Following [14], $E \subseteq \mathcal{L}$ is called a *fixed point* of a theory $T \subseteq \mathcal{L}$ if and only if $E = Th(T \cup \{\neg L\phi \mid \phi \in \mathcal{L} \perp E \text{ and } \phi \text{ is closed }\})$.

The assumption-based framework for non-monotonic modal logic is $\langle(\mathcal{L}, \mathcal{R}), Ab, \leq\rangle$ where

- $Ab = \{\neg L\phi \mid \phi \in \mathcal{L} \text{ and } \phi \text{ is closed }\}$,
- $\leq$ is defined by: $\quad \phi \leq \neg L\phi$, for any $\phi \in \mathcal{L}$.

In the assumption-based framework corresponding to non-monotonic modal logic, the notion of *attack* becomes:

- Given a theory $T$ and sets of assumptions $\Delta$ and $\mathcal{A}$,
  $\mathcal{A}$ *attacks* $\Delta$ if and only if
  $T \cup \mathcal{A} \vdash \phi$, for some $\neg L\phi \in \Delta$.

**Theorem 6.2** *$E$ is a fixed point of a theory $T$ if and only if $E$ is a stable extension of $T$ in the corresponding assumption-based framework.*

As elsewhere in this paper, the semantics can be improved by replacing stable extensions with weakly preferred extensions. For example, if the set $\mathcal{R}$ consists only of classical first-order logic and all instances of the necessitation rule, then weakly preferred extensions for the resulting framework, not only give the intuitively correct results for the theories $\{\neg L\phi \supset \phi\}$ and $\{L\phi\}$ of example 6.1, but also give the correct result, avoiding the anomalous extension, for the theory $\{L\phi \supset \phi\}$ of example 6.2.

# 7 Conclusions

The generalised framework for assumption-based reasoning demonstrates that different formalisms for non-monotonic reasoning are based upon similar principles. As a consequence, improvements made to the semantics of one formalism can be generalised and applied to other formalisms. We have illustrated this by arguing that admissible and preferred extensions are better than stable extensions and that $counterattacks_2$ is better than $counterattacks_1$. We first encountered this argument in the context of logic programming, but have investigated its generalisation and application to other formalisms for non-monotonic reasoning.

The generalised framework investigated in this paper is a variant of the argumentation framework presented by Dung. The two frameworks differ partly in their treatment of inconsistency and partly in the different levels of abstraction with which they treat the notions of assumptions, arguments and attacks. In preparing this paper we have investigated many variations of the definitions, most of which are mathematically equivalent. It is quite likely that further improvements can still be made. One particular matter which merits further consideration is the treatment of integrity constraints and whether they should participate in the generation of attacks, or should be confined to their present role in contributing only to inconsistencies.

In this paper we have limited our attention to matters of semantics. Proof procedures have been investigated in detail for the logic programming case and its extensions in other papers, and some of these are reported in the survey [13]. Proof procedures for the new semantics presented in this paper require further investigation. In particular, proof procedures generalising those developed for logic programming may also prove to be useful for other formalisms for non-monotonic reasoning.

## Acknowledgements

This research was supported by the Fujitsu Research Laboratories. The authors are grateful to Noboru Iwayama for having pointed out the mistake in definition 3.1.


## References


[1] A. Bondarenko, Abductive systems for non-monotonic reasoning. *Proc. 2nd Russian Conference on Logic Programming and Automated Reasoning* (A. Voronkov, ed.) Springer Verlag (1991)

[2] G. Brewka, Preferred subtheories: an extended logical framework for default reasoning. *Proc. 11th International Joint Conference on Artificial Intelligence,* Detroit, Mi (1989)

[3] A. Brogi, E. Lamma, P. Mello, P. Mancarella, Normal logic programs as open positive programs. *Proc. ICSLP '92* (1992)

[4] P. M. Dung, Negation as hypothesis: an abductive foundation for logic programming. *Proc. 8th International Conference on Logic Programming,* Paris, MIT Press (1991)

[5] P. M. Dung, P. Ruamviboonsuk, Well-founded reasoning with classical negation. *Proc. 1st International Workshop on Logic Programming and Nonmonotonic Reasoning*, Nerode, Marek and Subrahmanian eds., Washington DC (1991)

[6] P. M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning and logic programming. In preparation

[7] K. Eshghi, R. A. Kowalski, Abduction compared with negation as failure. *Proc. 6th International Conference on Logic Programming,* Lisbon, Portugal, MIT Press (1989)

[8] M. Gelfond, V. Lifschitz, The stable model semantics for logic programming. *Proc. 5th International Conference and Symposium on Logic Programming,* Washington, Seattle, MIT Press (1988)

[9] M. Gelfond, V. Lifschitz, Logic programs with classical negation. *Proc. 7th International Conference on Logic Programming,* Jerusalem, MIT Press (1990)

[10] A. C. Kakas, P. Mancarella, Generalized stable models: a semantics for abduction. *Proc. 9th ECAI,* Stockolm, (1990)

[11] A. C. Kakas, P. Mancarella, Stable theories for logic programs. *Proc. ISLP'91,* San Diego (1991)

[12] A. C. Kakas, P. Mancarella, Preferred extensions are partial stable models. *Journal of Logic Programming* 14(3,4) (1992)

[13] A. C. Kakas, R. A. Kowalski, F. Toni, Abductive logic programming. *Journal of Logic and Computation* 2(6) (1993)

[14] D. McDermott, Nonmonotonic logic II: nonmonotonic modal theories. *JACM* 29(1) (1982)

[15] R. Moore, Semantical considerations on non-monotonic logic. *Artificial Intelligence* 25 (1985)

[16] D. Poole, A logical framework for default reasoning. *Artificial Intelligence* 36 (1988)

[17] T. Przymusinski, Semantics of disjunctive logic programs and deductive databases. *Proc. DOOD '91* (1991)

[18] R. Reiter, A logic for default reasoning. *Artificial Intelligence* 13 (1980)

[19] D. Saccà, C. Zaniolo, Stable models and non-determinism for logic programs with negation. *Proc. ACM SIGMOD-SIGACT Symposium on Principles of Database Systems* (1990)

[20] A. Van Gelder, K.A. Ross, J.S. Schlipf, Unfounded sets and the well-founded semantics for general logic programs. *Proc. ACM SIGMOD-SIGACT, Symposium on Principles of Database Systems* (1988)