Relating defeasible and normal logic programming through transformation properties

Carlos Iván Chesñevar², Jürgen Dix²,*, Frieder Stolzenburg³, Guillermo Ricardo Simari
d
¹Department of Computer Science, Universidad Nacional del Sur Av. Alem 1253 - B8000CPB Bahía Blanca, Argentina
²Department of Computer Science, The University of Manchester, Oxford Road, Manchester M13 9PL, UK
³Institut für Informatik, Universität Koblenz-Landau Rheinau 1, 56075 Koblenz, Germany
⁴Department of Computer Science, Universidad Nacional del Sur, Av. Alem 1253 – B8000CPB Bahía Blanca, Argentina

Received 8 January 2001; received in revised form 9 November 2001; accepted 14 November 2001
Communication by G. Jäger

Abstract

This paper relates the Defeasible Logic Programming (DeLP) framework and its semantics SEM_{DeLP} to classical logic programming frameworks. In DeLP, we distinguish between two different sorts of rules: strict and defeasible rules. Negative literals ($\neg A$) in these rules are considered to represent classical negation. In contrast to this, in normal logic programming (NLP), there is only one kind of rules, but the meaning of negative literals (not $A$) is different: they represent a kind of negation as failure, and thereby introduce defeasibility. Various semantics have been defined for NLP, notably the well-founded semantics (WFS) (van Gelder et al., Proceedings of the Seventh Symposium on Principles of Database Systems, 1988, pp. 221–230; J. ACM 38 (3) (1991) 620) and the stable semantics Stable (Gelfond and Lifschitz, Fifth Conference on Logic Programming, MIT Press, Cambridge, MA, 1988, pp. 1070–1080; Proceedings of the Seventh International Conference on Logical Programming, Jerusalem, MIT Press, Cambridge, MA, 1991, pp. 579–597).

²This paper emerged while the first author was visiting the University of Koblenz in February 2000. It is part of the joint Argentine-German collaboration project DeReLoP (Dix et al., 1999). A preliminary version appeared at CACIC '2000 (Chesñevar et al., 2000a).

*Corresponding author.

E-mail addresses: cic@cs.uns.edu.ar (C.I. Chesñevar), dix@cs.man.ac.uk (J. Dix), stolzen@uni-koblenz.de (F. Stolzenburg), grs@cs.uns.edu.ar (G.R. Simari).
In this paper we consider the transformation properties for NLP introduced by Brass and Dix (J. Logic Programming 38(3) (1999) 167) and suitably adjusted for the DeLP framework. We show which transformation properties are satisfied, thereby identifying aspects in which NLP and DeLP differ. We contend that the transformation rules presented in this paper can help to gain a better understanding of the relationship of DeLP semantics with respect to more traditional logic programming approaches. As a byproduct, we obtain the result that DeLP is a proper extension of NLP. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Defeasible argumentation; Knowledge representation; Logic programming; Non-monotonic reasoning

1. Introduction and motivations

Defeasible logic programming (DeLP) [31,16,18] is a logic programming formalism which relies upon defeasible argumentation [28,9] for solving queries. DeLP combines strict rules, defined as in extended logic programming, and defeasible rules, of the form $A \leftarrow B$, indicating that reasons to believe in the antecedent $B$ provide reasons to believe in the consequent $A$. Solving a query $Q$ in DeLP gives rise to a proof $\mathcal{A}$ for $Q$ (written $\langle \mathcal{A}, Q \rangle$ for short) involving both strict and defeasible rules, called argument. In order to determine whether $Q$ is ultimately accepted as justified belief, a recursive analysis is performed which involves finding defeaters, i.e., arguments against accepting $\mathcal{A}$, which are better than $\mathcal{A}$ (according to a preference criterion). Since defeaters are arguments, a recursive procedure is to be carried out, in which defeaters, defeaters of defeaters, and so on, must be taken into account.

A development closely related to defeasible argumentation is so-called defeasible logic, initiated by Donald Nute [25]. Whereas the notion of defeat in defeasible argumentation is defined in terms of arguments, in defeasible logic it is defined between rules. Nute has developed a family of such logics, in which defeasible rules resemble Reiter’s defaults [29] (in the sense that they are one-directional). However, a special category of defeater rules is introduced, which are in a sense comparable to Pollock’s undercutting defeaters [26]. More details about Nute’s approach can be found in [25,28].

Logic programming has experienced considerable growth in the last decade, and several extensions have been developed and studied, such as normal logic programming (NLP) and extended logic programming (ENLP). For these formalizations different semantics have been developed, such as well-founded semantics and stable model semantics: we refer to [13,6,11] for an in-depth discussion of extensions of logic programming and their semantics. In contrast, DeLP has an “operational” semantics which is determined by the outcome of the dialectical process used for answering queries.

In [2,3,4], a number of transformation rules were introduced which allow one to “simplify” a normal logic program ($nlp$) $P$ to get its well-founded semantics WFS. The application of these rules leads to a new, simplified $nlp P'$ from which its WFS can be easily read off. In this paper, we will focus on finding similar transformation rules for DeLP, which can be used to simplify the knowledge encoded in a DeLP
program. In our analysis, we show that in DeLP a complete simplification of the original program cannot be achieved. However, our results suggest some connections between the semantics of classical approaches and logic programming with DeLP.

The paper is structured as follows: Section 2 introduces preliminary notions concerning NLP and DeLP. Section 3 introduces transformations for NLP. Section 4 shows how to adapt these transformations for DeLP, analyzing two classes of DeLP programs: DeLP_neg (Section 4.1) and DeLP_not (Section 4.2). Section 4.4 summarizes the relationships between NLP and DeLP, and the main results we have obtained. Finally, Section 5 discusses related work and concludes.

2. Preliminaries

In order to make this article self-contained, this section contains all the necessary definitions. Section 2.1 introduces normal logic programs, and Section 2.2 introduces the defeasible logic programming framework. We will focus our analysis on propositional logic programs because, following [19,23], program rules with variables can be viewed as “schemata” that represent their ground instances. Although there now exist powerful grounding techniques applied by various implementations (smodels, DLV), we believe that handling programs with free variables and computing appropriate substitutions (variable bindings) can often improve the performance of the system. Therefore, whenever suitable, we will also use the notion of most general unifiers (mgU) stemming from logic programming (see also Ref. [10]). Note that, in the following, an atom always may have variable or constant parameters only in this context. We do not consider the case with general functions here, because this would lead us to the problem of infinite programs and infinite argumentation lines (see also Proposition 2.25).

2.1. Normal Logic Programs (NLP)

Definition 2.1 (Normal logic program \( \mathcal{P} \)). A nlp \( \mathcal{P} \) is a finite set of normal program rules. A normal program rule has the form \( A \leftarrow L_1, \ldots, L_k \), where \( A \) is an atom and each \( L_i \) is an atom \( B \) or its negation not \( B \). If \( B = \{ L_1, \ldots, L_k \} \) is the body of a rule \( A \leftarrow B \), we also use the notation \( A \leftarrow B^+, \text{not } B^- \), where \( B^+ \) (resp. \( B^- \)) contains all the positive (resp. negative) body atoms in \( B \).

In NLP, atoms \( A \) and negated atoms not \( A \) are called literals. However, we must not confuse this notion with the notion of a literal introduced in Section 2.2 (Definition 2.3). In the sequel, we will speak of an atom and its negation, referring to an atom \( A \) and its default negation not \( A \). If \( B^+ = B^- = \emptyset \), we say that the rule is a fact and denote it by \( A \leftarrow \) (or just by \( A \)).

We will now introduce some concepts useful for describing what a semantics of a nlp is. Let \( \text{Prog}_\mathcal{P} \) be the set of all normal propositional programs with atoms from a signature \( \mathcal{L} \). By \( \mathcal{L}_\mathcal{P} \), we understand the signature of \( \mathcal{P} \), i.e., the set of atoms that occur in \( \mathcal{P} \). A (partial) interpretation based on a signature \( \mathcal{L} \) is a disjoint pair of sets
such that \( I_1 \cup I_2 \subseteq \mathcal{L} \). A partial interpretation is total if \( I_1 \cup I_2 = \mathcal{L} \). We may also view an interpretation \( \langle I_1, I_2 \rangle \) as the set of atoms and negated atoms \( I_1 \cup \text{not } I_2 \).

**Definition 2.2 (Semantics \( \text{SEM} \)).** A semantics \( \text{SEM} \) is a mapping which assigns to each logic program \( \mathcal{P} \) a set \( \text{SEM}(\mathcal{P}) \) of (partial) models of \( \mathcal{P} \), such that \( \text{SEM} \) is “instantiation invariant”, i.e., \( \text{SEM}(\mathcal{P}) = \text{SEM}(\text{ground}(\mathcal{P})) \), where \( \text{ground}(\mathcal{P}) \) denotes the Herbrand instantiation of \( \mathcal{P} \). A semantics \( \text{SEM} \) is called three-value based iff for each program \( \mathcal{P} \) the partial interpretation \( \text{SEM}(\mathcal{P}) \) is a three-valued model\(^1\) of \( \mathcal{P} \).

In Section 3, we will consider a particular three-valued semantics for the class \( \text{NLP} \) called the well-founded semantics \( \text{WFS} \), which can be computed by applying transformation rules on a \( \text{nlp} \mathcal{P} \).

### 2.2. Defeasible logic programs (DeLP)

The DeLP language \([31,16,18]\) is defined in terms of two disjoint sets of rules: a set of **strict rules** for representing strict (sound) knowledge, and a set of **defeasible rules** for representing tentative information. Rules will be defined using **literals**. A literal \( L \) is an atom \( p \) or a negated atom \( \sim p \), where the symbol “\( \sim \)” is called **strong negation**.

In addition, we will consider **default negation** with “not” here. We define formally:

**Definition 2.3 (Literal, assumption literal).** A literal \( L \) is an atom \( p \) or a negated atom \( \sim p \), where the symbol “\( \sim \)” represents **strong negation**. An assumption literal \( A \) has the form “not \( A \)”, where \( A \) is a literal.

**Definition 2.4 (Strict rules Head \( \leftarrow \) Body).** A **strict rule** is an ordered pair, conveniently denoted as \( \text{Head} \leftarrow \text{Body} \), the first member of which, \( \text{Head} \), is a literal, and the second member, \( \text{Body} \), is a finite set of literals, which may be (additionally) negated with “not” (default negation). A strict rule with the head \( L_0 \) and body \( \{L_1, \ldots, L_k\} \) can also be written as \( L_0 \leftarrow L_1, \ldots, L_k \). If the body is empty, it is written \( L \leftarrow \text{true} \), and it is called a **fact**. Facts may also be written as \( L \).

**Definition 2.5 (Defeasible rules Head \( \leftarrow\neg \) Body).** A **defeasible rule** is an ordered pair, conveniently denoted as \( \text{Head} \leftarrow\neg \text{Body} \), the first member of which, \( \text{Head} \), is a literal, and the second member, \( \text{Body} \), is a finite set of literals, which may be (additionally) negated with “not”. A defeasible rule with the head \( L_0 \) and body \( \{L_1, \ldots, L_k\} \) can also be written as \( L_0 \leftarrow\neg L_1, \ldots, L_k \), where \( k > 0 \).

A defeasible rule with an empty body (i.e. \( k = 0 \) in this case) is called a **presumption** \([16,17]\). Technically, it is possible to introduce presumptions into a framework for defeasible argumentation. However, this might lead to unintuitive results when comparing arguments by the definition of specificity we adopted in this paper (see Definition

\(^1\)We equip \( \leftarrow \) with the Kleene interpretation, where the implication \( \text{undef} \leftarrow \text{undef} \) is considered to be true.
2.16). For instance, two arguments solely based on presumptions are not comparable wrt specificity (according to Definition 2.16), although they should, if the set of presumptions used in one argument are a proper subset of the set of presumptions used in the other argument. Therefore, we will exclude presumptions from our object language.  

Syntactically, the symbol “— ” is all that distinguishes a defeasible rule from a strict rule. Defeasible rules account for tentative information that can be used if nothing can be argued against it, whereas strict rules are used to represent non-defeasible information.

In the sequel, atoms will be denoted with lowercase letters \((a, b, \ldots)\). The letter \(r\) (possibly indexed) will be used for denoting rule names. Literals will be denoted with capital letters \((A, B, \ldots)\), possibly indexed. Sets of atoms will be denoted as \(\mathcal{A}, \mathcal{B}, \ldots\), possibly indexed. Logic programs will be usually denoted as \(\mathcal{P}_1, \mathcal{P}_2, \text{ etc.}\)

**Definition 2.6** (Defeasible logic program \(\mathcal{P} = (\Pi, \Delta)\)). A defeasible logic program \(\text{dlp}\) is a finite set of strict and defeasible rules. If \(\mathcal{P}\) is a \(\text{dlp}\), we will distinguish in \(\mathcal{P}\) the subset \(\Pi\) of strict rules, and the subset \(\Delta\) of defeasible rules. When required, we will denote \(\mathcal{P}\) as \((\Pi, \Delta)\).

We will distinguish the class of all defeasible logic programs that use only strict (resp. default) negation, denoting them as \(\text{DeLP}_{\text{neg}}\) (resp. \(\text{DeLP}_{\text{not}}\)). Note that strong negation “\(\sim\)” is applied to atoms (also in rule heads), whereas default negation is applied to literals (possibly strongly negated). But default negation does not occur in heads of programs (see Definition 2.1). We will associate with every program \(\mathcal{P}\) a set of assumable facts of the form \(\text{assume}L\), for every literal \(L\) in \(\mathcal{P}\). Those literals will be given a special meaning in the argumentation framework. They will be used to define the semantics of default negation.

We will write \(\bar{\mathcal{P}}\) to denote the complement of a literal \(P\), defined as follows: \(\bar{\mathcal{P}} = \text{def} \sim P, \sim \mathcal{P} = \text{def} \bar{\mathcal{P}}, \text{ and assume} \mathcal{P} = \text{def} \bar{\bar{\mathcal{P}}}\).

We will define the notion of a defeasible derivation for a \(\text{dlp}\). In brief, it is a finite set of rules obtained by backward chaining from a literal \(Q\) as in a PROLOG program, using both strict and defeasible rules from the given \(\text{dlp}\ \mathcal{P}\). The symbol “\(\sim\)” is considered as part of the predicate when generating a defeasible derivation. The definition is similar to the one of SLDNF-derivation in Ref. [24], except that literals negated with “not” are associated with assumable facts.

**Definition 2.7** (Defeasible derivation). A defeasible derivation for a literal \(Q\) in a general \(\text{dlp}\ \mathcal{P}\) (possibly containing assumable facts) is a finite sequence of (instantiations of) rules in \(\mathcal{P}\). For this, we consider two sequences: a sequence \(G_i\) of goals i.e., sequences of sequences of literals, and a sequence \(r_i\) of rules for \(i \geq 0\) as follows:

1. \(G_0 = \{Q\}\). \(r_0\) is not defined.
2. Let \(G_i = \{Q_1, \ldots, Q_m, \ldots, Q_n\}\) with \(1 \leq m \leq n\).

\(^2\) For an in-depth analysis of presumptions with respect to \(\text{DeLP}\) the reader is referred to [17].
• If there is a strict rule \( r = (L_0 \leftarrow L_1, \ldots, L_k) \) in \( \mathcal{P} \), such that \( L_0 \) and \( Q_m \) have the most general unifier \( \sigma \), then \( G_{i+1} = [Q_1, \ldots, Q_{m-1}, L_1, \ldots, L_k, Q_{m+1}, \ldots, Q_n] \sigma \) and \( r_{i+1} = r \sigma \).

• If there is a defeasible rule \( r = (L_0 \nrightarrow L_1, \ldots, L_k) \) in \( \mathcal{P} \), such that \( L_0 \) and \( Q_m \) have the most general unifier \( \sigma \), then \( G_{i+1} = [Q_1, \ldots, Q_{m-1}, L_1, \ldots, L_k, Q_{m+1}, \ldots, Q_n] \sigma \) and \( r_{i+1} = r \sigma \).

• If \( Q_m \) has the form \( \text{not} \ L \) for some literal \( L \) (possibly negated with \( \neg \)) and the assumable fact \( r = \text{assume} \ L \) is in \( \mathcal{P} \), then \( G_{i+1} = [Q_1, \ldots, Q_{m-1}, Q_m+1, \ldots, Q_n] \) and \( r_{i+1} = r \).

The sequence of rules \( S = [r_1, \ldots, r_l] \) (for some suitable \( l > 0 \)) is called defeasible derivation for \( Q \) in \( \mathcal{P} \) if the corresponding goal (i.e., sequence of literals) \( G_i \) is empty. We say that \( Q \) can be defeasibly derived from \( \mathcal{P} \) and write \( \mathcal{P} \vdash Q \) in this case.

**Definition 2.8** (Contradictory set of rules). A set of rules \( \mathcal{S} \) is contradictory iff there is a defeasible derivation from \( \mathcal{S} \) for some literal \( P \) and its complement \( \neg P \), i.e., \( \mathcal{S} \vdash P \) and \( \mathcal{S} \vdash \neg P \).

Given a \( dlp \mathcal{P} \), we will always assume that the set \( \Pi \) of strict rules is non-contradictory (i.e., there is no literal \( P \) such that \( \Pi \vdash P \) and \( \Pi \vdash \neg P \)). If a contradictory set of strict rules were used in a \( dlp \), the same problems as in extended logic programming would appear. The corresponding analysis has been done elsewhere [20].

**Example 2.9.** Consider an engine the performance of which is determined by two switches \( sw_1 \) and \( sw_2 \). The switches regulate different features of the engine’s behavior, such as pumping system and working speed. We can model the engine behavior using a \( dlp \) \((\Pi, \Lambda)\), where

\[
\Pi = \{(sw_1 \leftarrow), (sw_2 \leftarrow), (heat \leftarrow), (\neg fuel.ok \leftarrow pump.clogged)\}
\]

(specifying that the two switches are on, there is heat, and whenever the pump gets clogged, fuel is not ok), and \( \Lambda \) models the possible behavior of the engine under different conditions (Fig. 1).

We now introduce the definition of argument in \( DeLP \). Basically, an argument for a literal \( Q \) is a defeasible derivation \( S = [r_1, \ldots, r_k] \) which is non-contradictory with respect to a given \( dlp \), and the defeasible information in \( S \) is minimal with respect to set inclusion.

**Definition 2.10** (Argument). Given a \( dlp \mathcal{P} = (\Pi, \Lambda) \), we will define \( \mathcal{H}_{\text{ass}} = \{ \text{assume} \ L | L \) is a literal in \( \mathcal{P} \} \). An argument \( \mathcal{A} \) for a query \( Q \), denoted \( (\mathcal{A}, Q) \), is defined as \( \mathcal{R}_A \cup \mathcal{H}_A \), where \( \mathcal{R}_A \) is a subset of ground instances of the defeasible rules of \( \mathcal{P} \) and \( \mathcal{H}_A \subseteq \mathcal{H}_{\text{ass}} \), such that:

1. there exists a defeasible derivation for \( Q \) from \( \Pi \cup \mathcal{A} \),
2. \( \Pi \cup \mathcal{A} \) is non-contradictory, and
3. \( \mathcal{A} \) is minimal with respect to set inclusion.
pump
fuel
ok
\rightarrow
sw_1

\text{(when \(sw_1\) is on, normally fuel is pumped properly)};

fuel
ok
\rightarrow
pump
fuel
ok

\text{(when fuel is pumped, normally fuel works ok)};

pump
oil
ok
\rightarrow
sw_2

\text{(when \(sw_2\) is on, normally oil is pumped)};

oil
ok
\rightarrow
pump
oil
ok

\text{(when oil is pumped, normally oil works ok)};

engine
ok
\rightarrow
fuel
ok, oil
ok

\text{(when there is fuel and oil, normally engine works ok)};

\sim
engine
ok
\rightarrow
fuel
ok, oil
ok, heat

\text{(when there is fuel, oil and heat, usually engine is not working ok)};

pump
\rightarrow
\text{clogged} pump
fuel
ok, low
speed

\text{(when fuel is pumped and speed is low, there are reasons to believe that the pump is clogged)};

low
speed
\rightarrow
sw_2

\text{(when \(sw_2\) is on, normally speed is low)};

\sim
low
speed
\rightarrow
sw_2, sw_1

\text{(when both \(sw_2\) and \(sw_1\) are on, speed is considered not to be low)}.

Fig. 1. Set \(\mathcal{A}\) (Example 2.9).

An argument \(\langle \mathcal{A}, Q \rangle\) is \textit{strict} iff \(\mathcal{A} = \emptyset\). An argument \(\langle \mathcal{A}_1, Q_1 \rangle\) is a \textit{sub-argument} of another argument \(\langle \mathcal{A}_2, Q_2 \rangle\) iff \(\mathcal{A}_1 \subseteq \mathcal{A}_2\). Given an argument \(\langle \mathcal{A}, Q \rangle\), we will also write \(\mathcal{H}(\langle \mathcal{A}, Q \rangle)\) to denote the set of assumption literals in \(\langle \mathcal{A}, Q \rangle\). We now introduce the auxiliary notion of \textit{immediate subargument}, which will be used later in the proofs of Propositions 4.12 and 4.24.

**Definition 2.11 (Immediate subarguments).** Let \(\langle \mathcal{A}, H \rangle\) be an argument, such that \(H \leftarrow P_1, \ldots, P_k\) is the last strict rule used in the defeasible derivation of \(H\) from \(\Pi \cup \mathcal{A}\). Clearly, in such a case there exist subsets \(\mathcal{A}_1, \ldots, \mathcal{A}_k\) of \(\mathcal{A}\), which are arguments for \(P_1, \ldots, P_k\). We will call \(\langle \mathcal{A}_1, P_1 \rangle, \ldots, \langle \mathcal{A}_k, P_k \rangle\) \textit{immediate subarguments} of \(\langle \mathcal{A}, H \rangle\).

**Example 2.12.** Consider the \(dlp\) \((\Pi, \Delta)\), with

\[
\Pi = \{(p \leftarrow q, \text{not } r), (w \leftarrow q, r), (s \leftarrow)\}
\]

\[
\Delta = \{(q \leftarrow < s), (r \leftarrow < s)\}
\]

It follows that \(\mathcal{A} = \{(q \leftarrow < s), (r \leftarrow < s)\}\) is an argument for \(w\), and \(\mathcal{B} = \{\text{assume } \sim r, (q \leftarrow < s)\}\) is an argument for \(p\). In the argument \(\langle \mathcal{B}, p \rangle\) the last strict rule used in the derivation of \(p\) is \(p \leftarrow q, \text{not } r\). Then \(\mathcal{B}' = \{q \leftarrow < s\}\) is an argument for \(q\), and it is an
immediate subargument of \( \langle \mathcal{A}, p \rangle \). In the argument \( \langle \mathcal{A}, w \rangle \) the last strict rule used in the derivation of \( w \) is \( w \leftarrow q, r \). Then \( \langle \mathcal{A}, q \rangle \) and \( \langle \mathcal{A}, r \rangle \) are immediate subarguments of \( \langle \mathcal{A}, w \rangle \).

**Example 2.13.** Consider Example 2.9. Then the set
\[
\mathcal{A} = \{ (pump\text{-}fuel\text{-}ok \leftarrow sw1), (pump\text{-}oil\text{-}ok \leftarrow sw2), \\
\text{fuel}\text{-}ok \leftarrow pump\text{-}fuel\text{-}ok, (oil\text{-}ok \leftarrow pump\text{-}oil\text{-}ok), \\
\text{engine}\text{-}ok \leftarrow fuel\text{-}ok, oil\text{-}ok \}
\]
is an argument for engine\text{-}ok. The set
\[
\mathcal{B} = \{ (pump\text{-}fuel\text{-}ok \leftarrow sw1), (low\text{-}speed \leftarrow sw2), \\
(pump\text{-}clogged \leftarrow pump\text{-}fuel\text{-}ok, low\text{-}speed) \}
\]
is an argument for \( \sim fuel\text{-}ok \). The set \( \mathcal{C} = \{ \sim low\text{-}speed \leftarrow sw2, sw1 \} \) is an argument for \( \sim low\text{-}speed \).

Given a dlp \( \mathcal{P} \), we will denote by \( \text{Args}(\mathcal{P}) \) the set of all possible arguments \( \langle \mathcal{A}, Q \rangle \) that can be built from \( \mathcal{P} \) wrt arbitrary queries \( Q \). We emphasize that this set consists of pairs \( \langle \mathcal{A}, Q \rangle \) and not just of arguments \( \mathcal{A} \) alone. This makes the condition \( \text{Args}(\mathcal{P}) = \text{Args}(\mathcal{P})' \) much stronger and is important for our Proposition 4.1 to hold.

The following definition captures the notion of conflict between two arguments.

**Definition 2.14 (Counterargument).** An argument \( \langle \mathcal{A}, Q_1 \rangle \) counter argues an argument \( \langle \mathcal{A}_2, Q_2 \rangle \) at a literal \( Q \) if there is a subargument \( \langle \mathcal{A}, Q \rangle \) of \( \langle \mathcal{A}_2, Q_2 \rangle \) such that \( \mathcal{N} \cup \{Q_1; Q\} \) is contradictory.

**Example 2.15.** Consider Example 2.13. Then \( \langle \mathcal{A}, \sim fuel\text{-}ok \rangle \) is a counterargument for \( \langle \mathcal{A}, \text{engine}\text{-}ok \rangle \), since there is a subargument \( \mathcal{A}' = \{ \text{fuel}\text{-}ok \leftarrow pump\text{-}fuel\text{-}ok, oil\text{-}ok \leftarrow pump\text{-}oil\text{-}ok, \text{engine}\text{-}ok \leftarrow fuel\text{-}ok, oil\text{-}ok \} \) for fuel\text{-}ok, such that \( \mathcal{N} \cup \{\text{fuel}\text{-}ok, \sim fuel\text{-}ok\} \) is contradictory.

Informally, a query \( Q \) will succeed if the supporting argument is not defeated; that argument becomes a justification. In order to establish that \( \mathcal{A} \) is a non-defeated argument, counterarguments that could be defeaters for \( \mathcal{A} \) are considered, i.e., counterarguments that are preferred to \( \mathcal{A} \) according to some criterion. DeLP considers a particular preference criterion called specificity [31,18] which favors an argument with greater information content and/or less use of defeasible rules. We introduce this concept more formally.

**Definition 2.16 (Specificity).** Given a dlp \( \mathcal{P} \), let \( \mathcal{P}_G \) denote the set of all rules with non-empty bodies. Let \( F \) denote the set of all possible literals that have a defeasible derivation in \( \mathcal{P} \).
An argument \( \langle \mathcal{A}_1, Q_1 \rangle \) is strictly more specific than an argument \( \langle \mathcal{A}_2, Q_2 \rangle \) (denoted \( \langle \mathcal{A}_1, Q_1 \rangle \succ \langle \mathcal{A}_2, Q_2 \rangle \)) if and only if:
1. For all \( H \subseteq F \); if \( \Pi_G \cup H \cup \mathcal{A}_1 \vdash Q_1 \) and \( \Pi_G \cup H \vdash Q_2 \), then \( \Pi_G \cup H \cup \mathcal{A}_2 \vdash Q_2 \).
2. There exists \( H' \subseteq F \) such that \( \Pi_G \cup H' \cup \mathcal{A}_1 \vdash Q_1 \) and \( \Pi_G \cup H' \vdash Q_2 \) and \( \Pi_G \cup H' \cup \mathcal{A}_2 \vdash Q_2 \).

**Example 2.17.** Consider the following \( dlp \mathcal{P} \):

\[
\mathcal{P} = \{ (p \leftarrow f_1, f_2), (\sim p \leftarrow f_1), (f_1 \leftarrow), (f_2 \leftarrow) \}
\]

Then the set of all literals derivable in \( \mathcal{P} \) is \( F = \{ p, \sim p, f_1, f_2 \} \). Consider the arguments \( \langle \mathcal{A}_1, p \rangle \) and \( \langle \mathcal{A}_2, \sim p \rangle \), with \( \mathcal{A}_1 = \{ p \leftarrow f_1, f_2 \} \) and \( \mathcal{A}_2 = \{ \sim p \leftarrow f_1 \} \). For every \( H \subseteq F \), condition 1 in Definition 2.16 holds. For \( H' = \{ f_1 \} \), condition 2 in Definition 2.16 holds. Hence \( \langle \mathcal{A}_1, p \rangle \) is strictly more specific than \( \langle \mathcal{A}_2, \sim p \rangle \).

**Definition 2.18 (Proper defeater, blocking defeater).** An argument \( \langle \mathcal{A}_1, Q_1 \rangle \) defeats \( \langle \mathcal{A}_2, Q_2 \rangle \) at a literal \( Q \) iff there exists a subargument \( \langle \mathcal{A}, Q \rangle \) of \( \langle \mathcal{A}_2, Q_2 \rangle \) such that \( \langle \mathcal{A}_1, Q_1 \rangle \) counter argues \( \langle \mathcal{A}, Q \rangle \) at \( Q \), and either:
1. \( \langle \mathcal{A}_1, Q_1 \rangle \) is strictly more specific than \( \langle \mathcal{A}, Q \rangle \). In this case \( \langle \mathcal{A}_1, Q_1 \rangle \) is called a proper defeater of \( \langle \mathcal{A}, Q \rangle \), or
2. Neither \( \langle \mathcal{A}_1, Q_1 \rangle \) is strictly more specific than \( \langle \mathcal{A}_2, Q_2 \rangle \), nor \( \langle \mathcal{A}_2, Q_2 \rangle \) is strictly more specific than \( \langle \mathcal{A}_1, Q_1 \rangle \). In this case \( \langle \mathcal{A}_1, Q_1 \rangle \) is a blocking defeater of \( \langle \mathcal{A}, Q \rangle \).

**Example 2.19.** Consider Examples 2.13 and 2.15. Then \( \langle \mathcal{B}, \sim fuel_{ok} \rangle \) is a proper defeater for \( \langle \mathcal{A}, engine_{ok} \rangle \), since it is more specific.

This conceptualization allows us to apply the notion of counterargumentation (Definition 2.14) and defeat (Definition 2.16) in a natural way when assumption literals are involved, as shown in the following example.

**Example 2.20.** Consider a \( dlp \mathcal{P} = (II, \Lambda) \), where

\[
II = \{ r \leftarrow s, t \leftarrow, q \leftarrow s \},
\]
\[
\Lambda = \{ p \leftarrow \text{not } q, r, q \leftarrow t \}
\]

Then \( \mathcal{A} = \{ p \leftarrow \text{not } q, r, \text{assume } \sim q \} \) is an argument for \( p \), which is counterargued by the argument \( \langle \{ q \leftarrow t \}, q \rangle \) as well as by the argument \( \langle \emptyset, q \rangle \).

Since defeaters are arguments, there may exist defeaters for the defeaters and so on. That prompts for a complete dialectical analysis to determine which arguments are ultimately defeated. Ultimately undefeated arguments will be marked as \( U \)-nodes, and the defeated ones as \( D \)-nodes. The formal definitions required for this process are as follows:
Definition 2.21 (Argumentation line). Let $\mathcal{P}$ be a dlp, and let $\langle \mathcal{A}, Q \rangle$ be an argument in $\mathcal{P}$. An argumentation line starting from $\langle \mathcal{A}, Q \rangle$, denoted $\lambda^{(\mathcal{A}, Q)}$ (or simply $\lambda$) is a possibly infinite sequence of arguments

$$\lambda^{(\mathcal{A}, Q)} = [\langle \mathcal{A}_0, Q_0 \rangle, \langle \mathcal{A}_1, Q_1 \rangle, \langle \mathcal{A}_2, Q_2 \rangle, \ldots, \langle \mathcal{A}_n, Q_n \rangle \ldots]$$

satisfying the following conditions:

1. If $\langle \mathcal{A}, Q \rangle$ has no defeaters, then $\lambda^{(\mathcal{A}, Q)} = [\langle \mathcal{A}, Q \rangle]$.
2. If $\langle \mathcal{A}, Q \rangle$ has a defeater $\langle \mathcal{B}, S \rangle$ in $\mathcal{P}$, then $\lambda^{(\mathcal{A}, Q)} = \langle \mathcal{A}, Q \rangle \circ \lambda^{(\mathcal{A}, S)}$.

We distinguish two sets in any argumentation line $\lambda$: the set of supporting arguments $\lambda_s = \{\langle \mathcal{A}_0, Q_0 \rangle, \langle \mathcal{A}_2, Q_2 \rangle, \langle \mathcal{A}_4, Q_4 \rangle, \ldots\}$ and the set of interfering arguments $\lambda_i = \{\langle \mathcal{A}_1, Q_1 \rangle, \langle \mathcal{A}_3, Q_3 \rangle, \langle \mathcal{A}_5, Q_5 \rangle, \ldots\}$.

Argumentation lines can be thought of as exchanges of arguments between two parties, a proponent and an opponent [30]. Dialectics imposes additional requirements on such an argument exchange to be considered rationally acceptable. In such a setting, fallacious reasoning (such as circular argumentation and falling into self-contradiction) is to be avoided. This can be done by requiring that all argumentation lines are acceptable [32]. An acceptable argumentation line starting with an argument $\langle \mathcal{A}_0, Q_0 \rangle$ constitutes an exchange of arguments which can be pursued until no more arguments can be introduced because of the dialectical constraints discussed above. These notions will be introduced in the following definitions.

Definition 2.22 (Contradictory set of arguments). Given a dlp $\mathcal{P} = (\Pi, \Delta)$, a set of arguments $S = \bigcup_{i=1}^{n} \{\langle \mathcal{A}_i, Q_i \rangle\}$ is contradictory wrt. $\mathcal{P}$ iff $\Pi \cup \bigcup_{i=1}^{n} \mathcal{A}_i$ is contradictory.

Definition 2.23 (Acceptable argumentation line). Let $\mathcal{P}$ be a dlp, and let $\lambda = [\langle \mathcal{A}_0, Q_0 \rangle, \langle \mathcal{A}_1, Q_1 \rangle, \ldots, \langle \mathcal{A}_n, Q_n \rangle, \ldots]$ be an argumentation line in $\mathcal{P}$. Let $\lambda' = [\langle \mathcal{A}_0, Q_0 \rangle, \langle \mathcal{A}_1, Q_1 \rangle, \ldots, \langle \mathcal{A}_k, Q_k \rangle, \ldots]$ be an initial segment of $\lambda$. The sequence $\lambda'$ is an acceptable argumentation line in $\mathcal{P}$ iff it is the longest initial segment in $\lambda$ satisfying the following conditions:

1. The sets $\lambda'_s$ and $\lambda'_i$ are each non-contradictory sets of arguments wrt $\mathcal{P}$.
2. No argument $\langle \mathcal{A}_i, Q_i \rangle$ in $\lambda'$ is a sub-argument of an earlier argument $\langle \mathcal{A}_j, Q_j \rangle$ of $\lambda'$ ($i < j$).
3. There is no subsequence of arguments $[\langle \mathcal{A}_{i-1}, Q_{i-1} \rangle, \langle \mathcal{A}_i, Q_i \rangle, \langle \mathcal{A}_{i+1}, Q_{i+1} \rangle]$ in $\lambda'$ such that $\langle \mathcal{A}_i, Q_i \rangle$ is a blocking defeater for $\langle \mathcal{A}_{i-1}, Q_{i-1} \rangle$ and $\langle \mathcal{A}_{i+1}, Q_{i+1} \rangle$ is a blocking defeater for $\langle \mathcal{A}_i, Q_i \rangle$.

The rationale for the conditions in Definition 2.23 can be better understood in a dialectical setting [32]. Condition 1 disallows the use of contradictory information on either side (proponent or opponent). Condition 2 eliminates the “circulus in demonstrando” fallacy (circular reasoning). Finally, condition 3 enforces the use of a stronger argument to defeat an argument which acts as a blocking defeater.
Example 2.24. Consider Example 2.9. The sequence

$$\lambda_1 = [\langle \mathcal{A}, \text{engine\_ok} \rangle, \langle \mathcal{B}, \sim \text{fuel\_ok} \rangle, \langle \mathcal{E}, \sim \text{low\_speed} \rangle]$$

is an acceptable argumentation line, whereas any sequence having the initial segment

$$\lambda_2 = [\langle \mathcal{A}, \text{engine\_ok} \rangle, \langle \mathcal{B}, \sim \text{fuel\_ok} \rangle, \langle \mathcal{D}, \text{fuel\_ok} \rangle]$$

with $\mathcal{D} = \{\text{pump\_fuel\_ok} \leftarrow \text{sw1, fuel\_ok} \leftarrow \text{pump\_fuel\_ok} \}$ is an argument line which is not acceptable, since the last argument defeats $\langle \mathcal{B}, \sim \text{fuel\_ok} \rangle$, but it is a subargument of a previous argument in $\lambda_2$ (viz. $\langle \mathcal{A}, \text{engine\_ok} \rangle$). Hence $\langle \mathcal{D}, \text{fuel\_ok} \rangle$ is deemed as a fallacious argument to be excluded from the dialectical analysis.

Proposition 2.25. Any acceptable argumentation line in a dlp $\mathcal{P}$ is finite.

Proof. Since $\mathcal{P}$ has no function symbols, and $\mathcal{P}$ is a finite set of program rules, the set of all possible arguments $\text{Args}(\mathcal{P})$ is necessarily finite. Hence the only way to get an infinite argumentation line $\hat{\lambda} = [\langle \mathcal{A}_0, Q_0 \rangle, \langle \mathcal{A}_1, Q_1 \rangle, \langle \mathcal{A}_2, Q_2 \rangle, \ldots, \langle \mathcal{A}_n, Q_n \rangle, \ldots]$ is by having the same argument twice in $\hat{\lambda}$, i.e., $\langle \mathcal{A}_i, Q_i \rangle = \langle \mathcal{A}_j, Q_j \rangle$, and hence $\mathcal{A}_i = \mathcal{A}_j$. But this cannot be the case in an acceptable argumentation line because of condition 2 in Definition 2.23. Therefore, any acceptable argumentation line $\hat{\lambda}$ is necessarily finite. \hfill $\Box$

Let $A^{(\mathcal{A}_0, Q_0)} = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ be the set of all acceptable argumentation lines starting with $\langle \mathcal{A}_0, Q_0 \rangle$ in a dlp $\mathcal{P}$. A tree structure can be built out of the elements of $A^{(\mathcal{A}_0, Q_0)}$, so that every path in the tree corresponds to some $\lambda_i \in A^{(\mathcal{A}_0, Q_0)}$. This structure will be called dialectical tree. Formally:

Definition 2.26 (Dialectical tree). Let $\mathcal{P}$ be a dlp, and let $\mathcal{A}_0$ be an argument for $Q_0$ in $\mathcal{P}$. A dialectical tree for $\langle \mathcal{A}_0, Q_0 \rangle$, denoted $\mathcal{T}^{(\mathcal{A}_0, Q_0)}$, is a tree structure defined as follows:

1. The root node of $\mathcal{T}^{(\mathcal{A}_0, Q_0)}$ is $\langle \mathcal{A}_0, Q_0 \rangle$.
2. $\langle \mathcal{B}', H' \rangle$ is an immediate child of $\langle \mathcal{B}, H \rangle$ iff there exists an acceptable argumentation line $\hat{\lambda}^{(\mathcal{A}_0, Q_0)} = [\langle \mathcal{A}_0, Q_0 \rangle, \langle \mathcal{A}_1, Q_1 \rangle, \ldots, \langle \mathcal{A}_n, Q_n \rangle]$ such that there are two elements $\langle \mathcal{A}_{i+1}, Q_{i+1} \rangle = \langle \mathcal{B}', H' \rangle$ and $\langle \mathcal{A}_i, Q_i \rangle = \langle \mathcal{B}, H \rangle$, for some $i = 0, \ldots, n - 1$.

Clearly, leaves in a dialectical tree correspond to undefeated arguments. Defeat among arguments in a dialectical tree can be propagated from the leaves up to the root, according to the marking procedure given in Definition 2.27.

Definition 2.27 (Marking of the dialectical tree). Let $\langle \mathcal{A}, Q \rangle$ be an argument and $\mathcal{T}^{(\mathcal{A}, Q)}$ its dialectical tree, then:

1. All the leaves in $\mathcal{T}^{(\mathcal{A}, Q)}$ are marked as $U$-nodes.
2. Let $\langle \mathcal{B}, H \rangle$ be an inner node of $\mathcal{T}^{(\mathcal{A}, Q)}$. Then $\langle \mathcal{B}, H \rangle$ will be a $U$-node iff each child of $\langle \mathcal{B}, H \rangle$ is a $D$-node. The node $\langle \mathcal{B}, H \rangle$ will be a $D$-node iff it has at least one child marked as $U$-node.
An argument $\mathcal{A}$ for a literal $Q$ which turns to be ultimately labelled as undefeated in $\mathcal{T}_{(\mathcal{A},Q)}$ is called a justification for $Q$.

**Definition 2.28 (Justification).** Let $\mathcal{A}$ be an argument for a literal $Q$, and let $\mathcal{T}_{(\mathcal{A},Q)}$ be its associated acceptable dialectical tree. The argument $\mathcal{A}$ for $Q$ will be a justification iff the root of $\mathcal{T}_{(\mathcal{A},Q)}$ is a $U$-node.

It can be shown [16] that for any dlp $\mathcal{P}$, strict arguments in $\mathcal{P}$ have no counterarguments, and therefore no defeaters. As a direct consequence of Definitions 2.26, 2.27 and 2.28, it follows that any strict argument $\mathcal{A}$ for a literal $Q$ will be a justification for $Q$: similar results hold for other argumentation systems, such as [35,27].

**Example 2.29.** Consider Example 2.9, and assume our main query is $\text{engine.ok}$. A argument $\langle \mathcal{A}, \text{engine.ok} \rangle$ can be built, which is defeated by the argument $\langle \mathcal{B}, \sim \text{fuel.ok} \rangle$ (as shown in Examples 2.13, 2.15 and 2.19). Hence, the argument $\langle \mathcal{A}, \text{engine.ok} \rangle$ will be provisionally rejected, since it is defeated. However, $\langle \mathcal{A}, \text{engine.ok} \rangle$ can be reinstated, since there exists a third argument $\mathcal{C} = \{\sim \text{low.speed} \leftarrow \text{sw1}, \text{sw2} \}$ for $\sim \text{low.speed}$ which in turn defeats $\langle \mathcal{B}, \sim \text{fuel.ok} \rangle$.

Hence, $\langle \mathcal{A}, \text{engine.ok} \rangle$ comes to be undefeated again, since the argument $\langle \mathcal{B}, \sim \text{fuel.ok} \rangle$ was defeated. But there is another defeater for $\langle \mathcal{A}, \text{engine.ok} \rangle$, the argument $\langle \mathcal{D}, \sim \text{engine.ok} \rangle$, where $\mathcal{D} = \{\text{pump.fuel.ok} \leftarrow \text{sw1}, \text{pump.oil.ok} \leftarrow \text{sw2}, \text{fuel.ok} \leftarrow \text{pump.fuel.ok}, \text{oil.ok} \leftarrow \text{pump.oil.ok}, \sim \text{engine.ok} \leftarrow \text{fuel.ok}, \text{oil.ok}, \text{heat} \}$. Hence $\langle \mathcal{D}, \text{engine.ok} \rangle$ is once again provisionally defeated.

Since there are no more arguments to consider, $\langle \mathcal{A}, \text{engine.ok} \rangle$ turns out to be ultimately defeated, so that we can conclude that the argument $\langle \mathcal{A}, \text{engine.ok} \rangle$ is not justified.

Fig. 2 shows the resulting dialectical tree, as well as its associated labelling.

A given query $Q$ can be associated with a particular answer set according to some criterion. Several criteria have been analyzed corresponding to different outcomes in the dialectical process. A possible criterion is specified in the following definition [16]:

![Fig. 2. Dialectical tree (Example 2.9).](image-url)
Definition 2.30 (Answers to a given query Q). Given a dlp \( P \), a query \( Q \) can be classified as a positive, negative, undecided or unknown answer as follows:

1. \( Q \) is a positive answer if there exists a justification \( \langle \mathcal{A}, Q \rangle \).
2. \( Q \) is a negative answer if for each argument \( \langle \mathcal{A}, Q \rangle \), in the dialectical tree \( T_{\langle \mathcal{A}, Q \rangle} \), there exists at least a proper defeater for \( \mathcal{A} \) marked as \( U \).
3. \( Q \) is an undecided answer if \( Q \) is not justified, and for each argument \( \langle \mathcal{A}, Q \rangle \), it is the case that \( T_{\langle \mathcal{A}, Q \rangle} \) has at least one blocking defeater marked as \( U \).
4. \( Q \) is an unknown answer if there is no argument for \( Q \).

Given a dlp \( P \), we call Positive(\( P \)), Negative(\( P \)), Undefined(\( P \)) and Unknown(\( P \)) the sets of positive, negative, undecided and unknown answers, respectively.

From the previous definition we can derive a three-valued semantics \( \text{SEM}_{\text{DLP}}(P) \) for a dlp \( P \), classifying literals in \( P \) as accepted, rejected or undefined as follows:

Definition 2.31 (\( \text{SEM}_{\text{DLP}} \)). For any dlp \( P \), we define \( \text{SEM}_{\text{DLP}}(P) = (\mathcal{P}_{\text{accepted}}, \mathcal{P}_{\text{rejected}}, \mathcal{P}_{\text{undefined}}) \), where

\[
\begin{align*}
\mathcal{P}_{\text{accepted}} &= \{ Q \mid Q \in \text{Justified}(P) \} \\
\mathcal{P}_{\text{rejected}} &= \{ Q \mid Q \in \text{Unknown}(P) \cup \text{Negative}(P) \} \\
\mathcal{P}_{\text{undefined}} &= \{ Q \mid Q \in \text{Undecided}(P) \}.
\end{align*}
\]

Example 2.32. Consider \( P \) as defined in Example 2.9, and consider the analysis performed in Example 2.29. Then \( \neg \text{engine.ok} \in \text{Negative}(P) \), \( \neg \text{engine.ok} \in \text{Positive}(P) \), \( \text{heat} \in \text{Positive}(P) \), and \( \neg \text{working.temperature_low} \in \text{Unknown}(P) \). Hence \( \{ \neg \text{engine.ok}, \text{heat} \} \subseteq \mathcal{P}_{\text{accepted}} \), and \( \text{engine.ok} \in \mathcal{P}_{\text{rejected}} \).

3. Transformations for NLP: classifying well-founded semantics

We are now considering logic programs containing default negation not. A program transformation is a relation \( \rightarrow \) between ground logic programs [2,4,5]. A semantics \( \text{SEM} \) allows a transformation \( \rightarrow \) iff \( \text{SEM}(P_1) = \text{SEM}(P_2) \), for all \( P_1 \) and \( P_2 \), such that \( P_1 \rightarrow P_2 \). In this case, we also say that the transformation \( \rightarrow \) holds wrt \( \text{SEM} \). Well-founded semantics for NLP can be elegantly characterized by a set of transformation rules [4], which reduce a given nlp program \( P \) into a simplified version \( P' \), from which the WFS can be easily read off.

Definition 3.1 (Transformation rules for WFS). Given a program \( P \in \text{Prog}_{\mathcal{P}} \), let \( \text{HEAD}(P) \) be the set of all head-atoms of \( P \), i.e., \( \text{HEAD}(P) = \{ H \mid H \leftarrow B^+, \text{not } \neg B^- \in P \} \). Let \( P_1 \) and \( P_2 \) be ground programs. The following transformation rules characterize WFS:

RED\(^+\) (Positive Reduction): Program \( P_2 \) results from program \( P_1 \) by RED\(^+\) (written \( P_1 \rightarrow^+ P_2 \)) iff there is a rule \( H \leftarrow B \) in \( P_1 \) and a negative literal \( \neg B \in B \) such
that there is no rule about \( B \) in \( P_1 \), i.e., \( B \notin \text{HEAD}(P_1) \), and \( P_2 = (P_1 \backslash \{ H \leftarrow B \}) \cup \{ H \leftarrow (B \backslash \{ not \ B \}) \} \).

**RED** (Negative Reduction): Program \( P_2 \) results from program \( P_1 \) by RED (written \( P_1 \mapsto N P_2 \)) iff there is a rule \( H \leftarrow B \) in \( P_1 \) and a negative literal \( \text{not} \ B \in B \) such that \( B \) appears as a fact in \( P_1 \), and \( P_2 = P_1 \backslash \{ H \leftarrow B \} \).

**SUB** (Deletion of non-minimal rules): Program \( P_2 \) results from program \( P_1 \) by SUB (written \( P_1 \mapsto M P_2 \)) iff there are rules \( H \leftarrow B \) and \( H \leftarrow B' \) in \( P_1 \) such that \( B \subset B' \) and \( P_2 = P_1 \backslash \{ H \leftarrow B' \} \).

**UNFOLD** (Unfolding): Program \( P_2 \) results from program \( P_1 \) by UNFOLD (written \( P_1 \mapsto U P_2 \)) iff there is a rule \( H \leftarrow B \) in \( P_1 \) and a positive literal \( B \in B \) such that \( P_2 = P_1 \backslash \{ H \leftarrow B \} \cup \{ H \leftarrow (B \backslash \{ B \}) \cup B' \} \backslash \{ B \leftarrow B' \} \).

**TAUT** (Deletion of Tautologies): Program \( P_2 \) results from program \( P_1 \) by TAUT (written \( P_1 \mapsto T P_2 \)) iff there is a rule \( H \leftarrow B \) in \( P_1 \) such that \( H \in B \) and \( P_2 = P_1 \backslash \{ H \leftarrow B \} \).

A program \( P' \) is a normal form of a program \( P \) wrt a transformation \( \mapsto \ast \) iff \( P \mapsto \ast P' \), where \( \mapsto \ast \) denotes the reflexive-transitive closure of \( \mapsto \), and \( P' \) is irreducible, i.e., there is no program \( P'' \) such that \( P' \mapsto \ast P'' \).

Let \( \mapsto \ast \) be the rewriting system consisting of the above five transformations, i.e., \( \mapsto \ast = \mapsto T \cup \mapsto U \cup \mapsto M \cup \mapsto p \cup \mapsto N \). Two distinctive features of this rewriting system [3] are that it is weakly terminating (i.e., each ground program \( P \) has a normal form \( P' \)), and confluent (i.e., given a program \( P \), by applying the transformations in any fair order, we eventually arrive at a normal form \( \text{norm}_{\text{WFS}}(P) \)). This normal form \( \text{norm}_{\text{WFS}}(P) \) is a residual program, consisting of rules without positive body atoms. For such a simplified program, its well-founded semantics can be easily read off as follows:

**Definition 3.2 (SEM\text{min}).** We define \( \text{SEM}_{\text{min}}(P) = (\mathcal{P}_{\text{true}}, \mathcal{P}_{\text{false}}, \mathcal{P}_{\text{undef}}) \) for any nlp \( P \), where

\[
\mathcal{P}_{\text{true}} = \{ H | H \leftarrow \in P \}
\]

\[
\mathcal{P}_{\text{false}} = \{ H | H \in L_P \backslash \text{HEAD}(P) \}
\]

\[
\mathcal{P}_{\text{undef}} = \{ H | H \in L_P \backslash (\mathcal{P}_{\text{true}} \cup \mathcal{P}_{\text{false}}) \}.
\]

To illustrate our transformations, we consider the following example taken from [14]:

**Example 3.3 (Computing WFS).** We consider the program \( P_1 \) (on the left) and reduce it as follows:

\[
\begin{align*}
p & \leftarrow \text{not} \ p & p & \leftarrow \text{not} \ p & p \\
q & \leftarrow \text{not} \ p & q & \leftarrow \text{not} \ p & s \leftarrow \text{not} \ q \\
s & \leftarrow \text{not} \ q & s & \leftarrow \text{not} \ q & s \leftarrow \text{not} \ q \\
r & \leftarrow \text{not} \ q & r & \leftarrow \text{not} \ q & q \leftarrow \text{not} \ q \\
r & \leftarrow q & r & \leftarrow q & r \leftarrow q
\end{align*}
\]
In the next step, we can apply **UNFOLD** to one of the two last rules to get:

\[
\begin{align*}
& p \\
\rightarrow_U & s \leftarrow \text{not } q \\
& q \leftarrow q \\
& r \leftarrow q
\end{align*}
\]

Now we can delete the resulting tautology by the application of **TAUT** and then use **Red**:

\[
\begin{align*}
& p \\
\rightarrow_T & s \leftarrow \text{not } q \quad \rightarrow p \quad s \\
& r \leftarrow q \quad r \leftarrow q
\end{align*}
\]

Finally, applying **UNFOLD** to the last one, we get to \( \text{norm}_{\text{WFS}}(\mathcal{P}_1) \):

\[
\begin{align*}
\rightarrow_U & p \\
\rightarrow & s
\end{align*}
\]

Thus, the well-founded semantics of \( \mathcal{P}_1 \) is

\[
\text{WFS}(\mathcal{P}_1) = \{ p, s, \text{not } q, \text{not } t, \text{not } r \}
\]

**Theorem 3.4** (Classifying WFS (Brass and Dix [4])). \( \text{WFS}(\mathcal{P}) = \text{SEM}_{\text{min}}(\text{norm}_{\text{WFS}}(\mathcal{P})) \).

### 4. Transformation properties in DeLP

As stated in the introduction, we want to analyze whether transformations for **NLP** as the ones described above also hold for a **DeLP** program. Such an analysis is very complicated for the whole class **DeLP**, where we have not only two sorts of rules, strict and defeasible rules, but also two different kinds of negation, \( \sim \) and \( \text{not} \). Adapting the transformation rules presented in Section 3 to this class of programs is a non-trivial task. In fact, even defining a semantics for general programs in **DeLP** is highly non-trivial and subject of ongoing research.

In our analysis, we will therefore focus first on **DeLP\textsubscript{neg}** (i.e., **DeLP** with strict negation “\( \sim \)”). As the transformations in Refs. [5,3] are defined with respect to a **NLP** setting, we will adapt them accordingly. Therefore, we extend our previous terminology to be applied to a **DeLP\textsubscript{neg}** program \( \mathcal{P} \) (thus HEAD(\( \mathcal{P} \)) will stand for all heads of rules in \( \mathcal{P} \), etc.), distinguishing strict rules from defeasible rules when needed. In Section 4.2, we will consider **DeLP\textsubscript{not}** (i.e., **DeLP** with default negation \( \text{not} \)). In that case, a similar analysis will be performed.

The following propositions provide ways of determining whether two dlps have the same semantics. These results will be used in the following sections.

**Proposition 4.1.** Let \( \mathcal{P} \) and \( \mathcal{P}' \) be two dlps. If \( \text{Args}(\mathcal{P}) = \text{Args}(\mathcal{P}') \), then \( \text{SEM}_{\text{DeLP}}(\mathcal{P}') = \text{SEM}_{\text{DeLP}}(\mathcal{P}) \).
Proof. This is a direct consequence of Definition 2.31, since the semantics of DeLP is entirely determined by relationships among arguments.

The converse does not hold, as shown in the following example.

Example 4.2. Let $\mathcal{P}_1 = \{p \leftarrow q, p \leftarrow r, q \leftarrow r\}$, and let $\mathcal{P}_2 = \{p \leftarrow q, q \leftarrow, r \leftarrow\}$. Clearly, $\text{SEM}_{\text{DeLP}}(\mathcal{P}_1) = \text{SEM}_{\text{DeLP}}(\mathcal{P}_2)$, since $\{p, q, r\} = \mathcal{P}_{1,\text{accepted}} = \mathcal{P}_{2,\text{accepted}}$. However $\text{ Args}(\mathcal{P}_1) \neq \text{ Args}(\mathcal{P}_2)$ (since $\{p \leftarrow r\}, p$) is an argument in $\mathcal{P}_1$ but not in $\mathcal{P}_2$.

Definition 4.3 (Isomorphic dialectical trees). Given two arguments $\langle A_1, Q_1 \rangle$ and $\langle A_2, Q_2 \rangle$, their associated dialectical trees $\mathcal{T}_{\langle A_1, Q_1 \rangle}$ and $\mathcal{T}_{\langle A_2, Q_2 \rangle}$ are isomorphic iff

1. $Q_1 = Q_2$, and both $\langle A_1, Q_1 \rangle$ and $\langle A_2, Q_2 \rangle$ have no defeaters, or
2. $\mathcal{T}_{\langle A_1, Q_1 \rangle}$ has $\mathcal{T}_1, \ldots, \mathcal{T}_k$ as immediate subtrees, and $\mathcal{T}_{\langle A_2, Q_2 \rangle}$ has $\mathcal{T}_1', \ldots, \mathcal{T}_k'$ as immediate subtrees, and there exists a one-to-one correspondence $f: \{\mathcal{T}_1, \ldots, \mathcal{T}_k\} \mapsto \{\mathcal{T}_1', \ldots, \mathcal{T}_k'\}$, such that
   (a) $\mathcal{T}_i$ and $f(\mathcal{T}_i)$ are isomorphic, $i = 1, \ldots, k$, and
   (b) The root of $\mathcal{T}_i$ is a proper (resp. blocking) defeater for $\langle A_1, Q_1 \rangle$ and the root of $f(\mathcal{T}_i)$ is a proper (resp. blocking) defeater for $\langle A_2, Q_2 \rangle$, for $i = 1, \ldots, k$.

Proposition 4.4. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two DeLP programs, such that $\mathcal{T}_{\langle A_1, Q_1 \rangle}$ is the associated dialectical tree for an argument $\langle A_1, Q_1 \rangle$ in $\mathcal{P}_1$, and $\mathcal{T}_{\langle A_2, Q_2 \rangle}$ is the associated dialectical tree for an argument $\langle A_2, Q_2 \rangle$ in $\mathcal{P}_2$. If $\mathcal{T}_{\langle A_1, Q_1 \rangle}$ and $\mathcal{T}_{\langle A_2, Q_2 \rangle}$ are isomorphic, then $Q_1 \in \mathcal{P}_{1,\text{accepted}}$ (resp. $\mathcal{P}_{1,\text{rejected}}$, $\mathcal{P}_{1,\text{undef}}$) iff $Q_2 \in \mathcal{P}_{2,\text{accepted}}$ (resp. $\mathcal{P}_{2,\text{rejected}}$, $\mathcal{P}_{2,\text{undef}}$).

Proof. This proposition is a direct consequence of the definition of marking of a dialectical tree (Definition 2.27).

Corollary 4.5. Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two DeLP programs, such that $\text{HEAD}(\mathcal{P}_1) = \text{HEAD}(\mathcal{P}_2)$. Suppose that for any literal $Q$ in $\text{HEAD}(\mathcal{P}_1)$, there exists a dialectical tree $\mathcal{T}_{\langle A_1, Q \rangle}$ in $\mathcal{P}_1$ iff there exists an isomorphic dialectical tree $\mathcal{T}_{\langle A_2, Q \rangle}$ in $\mathcal{P}_2$. Then $\text{SEM}_{\text{DeLP}}(\mathcal{P}_1) = \text{SEM}_{\text{DeLP}}(\mathcal{P}_2)$.

4.1. Transformation properties in DeLP

Below we will introduce tentative extensions to DeLP of the previous transformation rules. The distinguishing features of the transformation rules are discussed next. For each transformation, $\mathcal{P}_1$ and $\mathcal{P}_2$ denote ground dlps. Some transformation rules have special requirements which appear in boldface.

RED$_{\text{neg}}^+$: Program $\mathcal{P}_2$ will result from program $\mathcal{P}_1$ by RED$_{\text{neg}}^+$ (written $\mathcal{P}_1 \mapsto_{\text{P}\text{-neg}}^+ \mathcal{P}_2$) iff there is a rule $H \leftarrow \mathcal{B}$ in $\mathcal{P}_1$ and a negative literal $\sim \mathcal{B} \in \mathcal{B}$ such that there is no rule about $\mathcal{B}$ in $\mathcal{P}_1$, i.e., $\mathcal{B} \not\in \text{HEAD}(\mathcal{P}_1)$, and $\mathcal{P}_2 = (\mathcal{P}_1 \setminus \{H \leftarrow \mathcal{B}\}) \cup \{H \leftarrow (\mathcal{B} \setminus \{\sim \mathcal{B}\})\}$.

RED$_{\text{neg}}$: Program $\mathcal{P}_2$ will result from program $\mathcal{P}_1$ by RED$_{\text{neg}}$ (written $\mathcal{P}_1 \mapsto_{\text{M}\text{-neg}} \mathcal{P}_2$) iff there is a strict rule $H \leftarrow \mathcal{B}$ (or defeasible rule $H \leftarrow \mathcal{B}$) in $\mathcal{P}_1$ and a negative
Let $P_2$ will result from program $P_1$ by $\text{SUB}_\text{neg}$ (written $P_1 \rightarrow \text{SUB}_\text{neg} P_2$) iff there are strict rules $H \leftarrow B$ and $H \leftarrow B'$ in $P_1$ such that $B \subset B'$ and $P_2 = P_1 \setminus \{H \leftarrow B\}$. The rule $H \leftarrow B_2$ is called non-minimal rule wrt $H \leftarrow B_1$.

$\text{UNFOLD}_\text{neg}$: Suppose program $P_1$ contains a strict rule $H \leftarrow B$ such that there is no defeasible rule in $P_1$ with head $H$.

Then program $P_2$ will result from program $P_1$ by $\text{UNFOLD}_\text{neg}$ (written $P_1 \rightarrow \text{UNFOLD}_\text{neg} P_2$) iff there is a literal\(^3\) which does not appear as head of a defeasible rule in $P_1$, such that $P_2 = P_1 \setminus \{H \leftarrow B\} \cup \{H \leftarrow (B \setminus \{B\}) \cup B' | B \leftarrow B' \in P_1\}$. The clause $H \leftarrow B$ is said to be $\text{UNFOLD}_\text{neg}$-related with each $B \leftarrow B_i \in P_1$ (for $i = 1, \ldots, n$).

$\text{TAUT}_\text{neg}$: Program $P_2$ will result from program $P_1$ by $\text{TAUT}_\text{neg}$ (written $P_1 \rightarrow \text{TAUT}_\text{neg} P_2$) iff there is a strict rule $H \leftarrow B \in P_1$ (or defeasible rule $H \leftarrow B$) such that $H \in B$ and $P_2 = P_1 \setminus \{H \leftarrow B\}$ (or $P_2 = P_1 \setminus \{H \leftarrow B\}$).

First, we consider $\text{RED}^+$. This transformation rule does not hold for strict negation, i.e., equivalence is not preserved for $\text{DeLP}_\text{neg}$ programs, as the following Example 4.6.

Note that whereas $\text{RED}^+$ captures the idea that not $A$ trivially holds whenever $A$ cannot be derived (and for that reason not $A$ can be deleted), the same principle cannot be applied to $\sim A$, which holds whenever there is a derivation for $\sim A$. Note also that strict negation in contrast to default negation may appear in heads (as in the following example).

**Example 4.6.** Consider the following $\text{DeLP}_\text{neg}$ program: $\Pi = \{(p \leftarrow \sim s), (\sim s \leftarrow t), (q_1 \leftarrow), (q_2 \leftarrow\}\} \text{ and } A = \{(q_1 \leftarrow q_1), (\sim t \leftarrow q_1, q_2\}\}$. Here $p$ is not justified from $\mathcal{P}$ (since the argument $\mathcal{A}_1 = \{t \leftarrow q_1\}$ for $p$ is defeated by the argument $\mathcal{A}_2 = \{\sim t \leftarrow q_1, q_2\}$ for $\sim t$). If we considered $\mathcal{P}' = \text{RED}_\text{neg}(\mathcal{P})$ we would get $p$ as a fact (because there is no rule for $s$ in $\mathcal{P}$), so $p$ would be justified in $\mathcal{P}'$.

Let us now consider $\text{RED}^-$. This transformation rule holds for both defeasible and strict rules in a $\text{DeLP}_\text{neg}$ program $\mathcal{P}$, as shown in Proposition 4.7.

**Proposition 4.7.** Let $\mathcal{P}$ be a $\text{DeLP}_\text{neg}$ program. Let $\mathcal{P}'$ be the resulting program of applying $\text{RED}^-\text{neg}$, i.e., $\mathcal{P} \rightarrow \text{RED}^-\text{neg} \mathcal{P}'$. Then $\text{SEM}_{\text{DeLP}}(\mathcal{P}') = \text{SEM}_{\text{DeLP}}(\mathcal{P})$.

**Proof.** Let $\mathcal{P}$ be a $\text{DeLP}_\text{neg}$ program, and let $(A \leftarrow) \in \mathcal{P}$. Furthermore, let $r = P \leftarrow Q_1, \ldots, Q_n$ (resp. $P \leftarrow Q_1, \ldots, Q_n$) be a rule in $\mathcal{P}$, such that $\sim A \equiv Q_i$, for some $i$. Then $r$ cannot be used in any defeasible derivation corresponding to an argument in $\mathcal{P}$ (since if $r$ is used, then both $\sim A$ and $A$ follow from $\Pi \cup \mathcal{A}$, contradicting the definition of argument). Therefore, every argument that can be built from $\mathcal{P}$ can also be built from $\mathcal{P}' = \mathcal{P} \setminus \{r\}$. Thus $\text{Arg}_\mathcal{P}(\mathcal{P}) = \text{Arg}_\mathcal{P}(\mathcal{P}')$, and therefore $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$. \[\square\]

\(^3\)Note that we do not distinguish between atoms and their negations because negated literals are treated as new predicate names.
Let us now consider $\text{SUB}_{\text{neg}}$. This transformation holds for strict rules, as shown in Proposition 4.9. It does not hold in $\text{DeLP}_{\text{neg}}$ for defeasible rules (since having more literals in the body gives more specific information), as shown in Example 4.8.

**Example 4.8.** Let $\mathcal{P} = (II, \Lambda)$, where $II = \{q_1, q_2\}$ and $\Lambda = \{(p \leftarrow < q_1, q_2), (p \leftarrow < q_1)\}$. The argument $\mathcal{A} = \{(p \leftarrow < q_1, q_2)\}$ for $p$ is strictly more specific than $\mathcal{B} = \{(\sim p \leftarrow < q_2)\}$ for $\sim p$. However, if we consider $\mathcal{P}' = \mathcal{P} \setminus \{(p \leftarrow < q_1, q_2)\}$, then we get two arguments which block each other ($\mathcal{A} = \{(p \leftarrow < q_1)\}$ for $p$ and $\mathcal{B} = \{(\sim p \leftarrow < q_2)\}$ for $\sim p$).

**Proposition 4.9.** Let $\mathcal{P}$ be a $\text{DeLP}_{\text{neg}}$ program. Let $\mathcal{P}'$ be the program resulting from applying $\text{SUB}_{\text{neg}}$, i.e., $\mathcal{P} \rightarrow_{\text{Mneg}} \mathcal{P}'$. Then $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$.

**Proof.** Clearly, $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P} \setminus \{r \mid r \text{ non-minimal rule}\})$, as the following line of reasoning shows. Let $r = p \leftarrow Q_1, \ldots, Q_k$ be a non-minimal rule in $\mathcal{P}$, and assume there is an argument $\mathcal{A}$ for some literal $H$ in which $r$ is part of the defeasible derivation for $H$. From the definition of defeasible derivation, for each literal $Q_1, \ldots, Q_k$ there is an argument $\langle B_1, Q_1 \rangle, \ldots, \langle B_k, Q_k \rangle$, such that $\bigcup_{i=1}^{k} B_i \subseteq \mathcal{A}$. Since $r$ is a non-minimal rule, there exists $r' = p \leftarrow Q_1, \ldots, Q_j$ such that for each literal $Q_i$ ($i = 1, \ldots, j$) there are arguments $\langle B_1, Q_1 \rangle, \ldots, \langle B_k, Q_k \rangle$. Hence by replacing $r$ by $r'$, we get either the same set $\mathcal{A}$ as an argument for $H$, or a proper subset $\mathcal{A}' \subseteq \mathcal{A}$ must be an argument for $H$. This means that $\mathcal{A}$ is not an argument according to Definition 2.10, because it does not satisfy condition 3. In any case, the rule $r$ can be removed from $\mathcal{P}$, without affecting the arguments that can be obtained from $\mathcal{P}$. Therefore, $\text{Args}(\mathcal{P}) = \text{Args}(\mathcal{P}')$ ($\mathcal{P}' = \mathcal{P} \setminus \{r\}$). Hence $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$. $\square$

Let us now consider $\text{UNFOLD}_{\text{neg}}$. As indicated in its definition, this property is only defined for a certain class of strict rules. It does not hold for defeasible rules, as shown in Example 4.10. It does not hold for strict rules in general either: we imposed the additional condition that no defeasible rule has the same head as the literal which is being removed when applying “unfolding”. The reason for doing so is shown in Example 4.11.

**Example 4.10 (UNFOLD does not hold for defeasible rules).** Consider the following example

<table>
<thead>
<tr>
<th>II</th>
<th>$\Lambda$</th>
</tr>
</thead>
<tbody>
<tr>
<td>has_feathers ← flies ← bird</td>
<td></td>
</tr>
<tr>
<td>has_beak ← ¬flies ← bird, wounded</td>
<td></td>
</tr>
<tr>
<td>wounded ← bird ← has_feathers, has_beak</td>
<td></td>
</tr>
</tbody>
</table>

In $\mathcal{P}$, there is an argument $\mathcal{A} = \{¬flies ← bird, wounded\}, (bird ← has_feathers, has_beak\}$ for $¬flies$ which is strictly more specific than $\mathcal{A}_2 = \{flies ← bird, bird ← has_feathers, has_beak\}$ for $flies$. In this case, the first argument is a justification. However, if $\text{UNFOLD}_{\text{neg}}$ is applied to defeasible rules, we get
\( \mathcal{P}' = (\Pi, \Delta') \), with \( \Delta' = \{(\text{flies} \leftarrow \text{has feathers, has beak}), (\sim\text{flies} \leftarrow \text{bird, wounded}), (\text{bird} \leftarrow \text{has feathers, has beak})\} \). In \( \mathcal{P}' \) we have two conflicting arguments, \( \mathcal{A}_1 = \{(\sim\text{flies} \leftarrow \text{bird, wounded}), (\text{bird} \leftarrow \text{has feathers, has beak})\} \) for \( \sim\text{flies} \) and \( \mathcal{A}_2 = \{(\text{flies} \leftarrow \text{has feathers, has beak})\} \) for \( \text{flies} \). In this case, neither of them is strictly more specific than the other.

**Example 4.11.** Let \( \mathcal{P} = (\Pi, \Delta) \) be a dlp, where \( \Pi = \{(p \leftarrow q, s), (q \leftarrow f_1), (q \leftarrow f_2), (s \leftarrow)\} \), and \( \Delta = \{q \leftarrow s\} \). If we could apply UNFOLD\_neg on rule \( (p \leftarrow q, s) \) wrt the literal \( q \), we would get the program \( \mathcal{P}' = \mathcal{P} \setminus \{(p \leftarrow q, s)\} \cup \{(p \leftarrow f_1, s), (p \leftarrow f_2, s)\} \). But \( \mathcal{A}_1 = \{q \leftarrow s\} \) is an argument for \( p \) in \( \mathcal{P} \), but it does not exist in \( \mathcal{P}' \).

In order to simplify the analysis of UNFOLD\_neg, we will define a special transformation UNFOLD\_neg\_r corresponding to UNFOLD\_neg applied to a particular UNFOLD\_neg-related rule \( r \).

**Definition 4.12** (Transformation UNFOLD\_neg\_r). Suppose program \( \mathcal{P}_1 \) contains a strict rule \( H \leftarrow B \) such that there is no defeasible rule in \( \mathcal{P}_1 \) with head \( H \).

Then program \( \mathcal{P}_2 \) will result from program \( \mathcal{P}_1 \) by UNFOLD\_neg\_r (written \( \mathcal{P}_1 \leftarrow \mathcal{P}_2 \)) if there is a positive literal \( B \in \mathcal{B} \) which does not appear as head of a defeasible rule in \( \mathcal{P}_1 \), such that \( \mathcal{P}_2 = \mathcal{P}_1 \setminus \{H \leftarrow B\} \cup \{H \leftarrow ((\mathcal{B} \setminus \{B\}) \cup \mathcal{B}')\} \) s.t. \( r_i = B \leftarrow \mathcal{B}' \in \mathcal{P}_1 \). (Such \( r_i \) are called UNFOLD\_neg-related.)

**Proposition 4.13.** Let \( \mathcal{P}_i \) be a DeLP\_neg program which contains a strict rule \( r = H \leftarrow B \), such that \( r_1, r_2, \ldots, r_k \) are all those rules in \( \mathcal{P}_1 \) that are UNFOLD\_neg-related to \( r \). Consider the sequence of programs \( \mathcal{P} = \mathcal{P}_1 \leftarrow \mathcal{P}_2 \leftarrow \mathcal{P}_3 \leftarrow \ldots \leftarrow \mathcal{P}_k \leftarrow \mathcal{P}' \).

Then \( \mathcal{P} \leftarrow \mathcal{P}' \) wrt rule \( r \).

**Proof.** Direct consequence of Definition 4.12 and the definition of UNFOLD\_neg.

We present next a particular property of immediate subarguments in DeLP\_neg, which will allow us to show that the transformation \( \leftarrow \mathcal{P}' \) preserves semantics when applied to a given DeLP\_neg program.

**Proposition 4.14.** Let \( \langle \mathcal{A}, H \rangle \) be an argument in DeLP\_neg, such that the last rule used in the derivation is the strict rule \( H \leftarrow P_1, \ldots, P_k \). Then all immediate subarguments \( \langle \mathcal{A}_1, P_1 \rangle, \ldots, \langle \mathcal{A}_k, P_k \rangle \) are such that \( \mathcal{A}_i = \mathcal{A} \), for \( i = 1, \ldots, k \).

**Proof.** Since \( \langle \mathcal{A}, H \rangle \) is an argument, then \( \Pi \cup \mathcal{A} \vdash H \), such that there exists a defeasible derivation \( S = [r_1, \ldots, r_k] \), where \( r_1 = H \leftarrow P_1, \ldots, P_k \). Clearly, the sequence \( S' = [r_2, \ldots, r_k] \) provides a defeasible derivation for every element of the sequence of goals \( G = [P_1, \ldots, P_k] \), using the same set \( \mathcal{A} \) of defeasible information as in \( S \). In particular, \( \Pi \cup \mathcal{A} \vdash P_i \), \( \forall i = 1, \ldots, k \), such that \( \mathcal{A} \) is minimal and non-contradictory. Thus \( \mathcal{A} \) is an argument for \( P_i \), \( \forall i = 1, \ldots, k \).
Proposition 4.15. Let $\mathcal{P}_1$ be a DeLP$^\neg$ program, and let $\mathcal{P}_2$ be the program resulting from applying $\neg\neg$ wrt some rule $r_i$.

Let $\langle A, H \rangle$ be an argument in $\mathcal{P}_1$ affected by the application of $\neg\neg$. Then $\langle A, H \rangle$ is also an argument in $\mathcal{P}_2$, and $\text{Args}(\mathcal{P}_1) = \text{Args}(\mathcal{P}_2)$.

Proof. Let $\mathcal{P}_1 = (\Pi, \Delta)$ be a DeLP$^\neg$ program. Let $\langle A, Q \rangle$ be an argument in $\mathcal{P}_1$. We can assume that (i) a strict rule $r = H \leftarrow B$ is used in the defeasible derivation of $Q$ from $\Pi \cup A$, and (ii) $r$ is UNFOLD-related to other rule $r_i$ (otherwise $\text{Args}(\mathcal{P}_1) = \text{Args}(\mathcal{P}_2)$, and the proposition holds trivially).

Since rule $r$ was applied in the defeasible derivation of $Q$ from $\Pi \cup A$, there exists an argument $\langle A', H \rangle$ which is a subargument of $\langle A, Q \rangle$, such that the last rule used in the defeasible derivation of $\langle A', H \rangle$ is $r$. The strict rule $r$ can be written as

$$r = H \leftarrow B, L_1, \ldots, L_k$$

From Proposition 4.14, we get that $\langle A, B \rangle, \langle A, L_1 \rangle, \ldots, \langle A, L_k \rangle$ are immediate subarguments of $\langle A', H \rangle$.

Consider $r_i = B \leftarrow B$, which is the last rule used in the defeasible derivation of $\langle A, B \rangle$, such that $r$ is UNFOLD-related to $r_i$. Since $r_i$ is a strict rule, it will have the form

$$r_i = B \leftarrow P_1, \ldots, P_m.$$  

From Proposition 4.14, we get that $\langle A, P_1 \rangle, \langle A, P_2 \rangle, \ldots, \langle A, P_m \rangle$ are immediate subarguments of $\langle A, B \rangle$. Thus, argument $\langle A, H \rangle$ in $\mathcal{P}_1$ is such that $\langle A, P_1 \rangle, \ldots, \langle A, P_m \rangle$ and $\langle A, L_1 \rangle, \ldots, \langle A, L_k \rangle$ are also arguments in $\mathcal{P}_1$.

Assume we apply $\neg\neg$ to $\mathcal{P}_1$, resulting in a new DeLP$^\neg$ program $\mathcal{P}_2$. From Definition 4.12, we have:

$$\mathcal{P}_2 = \mathcal{P}_1 \setminus \{H \leftarrow B\} \cup \{H \leftarrow ((\mathcal{B} \setminus \{B\}) \cup \mathcal{B})\}$$

In this case we get $\mathcal{P}_2 = \mathcal{P}_1 \setminus \{r\} \cup \{r'\}$, where $r'$ is the rule

$$r' = H \leftarrow L_1, \ldots, L_k, P_1, \ldots, P_m.$$  

Clearly, $\langle A, L_i \rangle, i = 1, \ldots, k$ and $\langle A, P_i \rangle, i = 1, \ldots, m$ are also arguments in $\mathcal{P}_2$, and in particular $\langle A, H \rangle$ is also an argument in $\mathcal{P}_2$. Note that no new argument other than $\langle A, H \rangle$ is generated in $\mathcal{P}_2$, since the subarguments of $\langle A, H \rangle$ in $\mathcal{P}_1$ and $\langle A, H \rangle$ in $\mathcal{P}_2$ are the same. Thus, $\text{Args}(\mathcal{P}_1) = \text{Args}(\mathcal{P}_2)$. □

Corollary 4.16. Let $\mathcal{P}$ be a DeLP$^\neg$ program, and let $\mathcal{P}'$ be the program resulting from applying UNFOLD$^\neg$ wrt some rule $r$ in $\mathcal{P}$. Then $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$.

Proof. Follows directly from Proposition 4.13 by repeated application of $\neg\neg$, for each $r_i$ which is UNFOLD-related with $r$. □

Let us now consider tautology elimination.
Proposition 4.17. Let $\mathcal{P}$ be a DeLP\textsubscript{not} program, and $\mathcal{P}'$ the program resulting from applying $\text{TAUT}\textsubscript{not}$ to $\mathcal{P}$, i.e., $\mathcal{P} \rightarrow_{\text{neg}} \mathcal{P}'$. Then $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$.

Proof. Let $\langle A, Q \rangle$ be an argument in $\text{Args}(\mathcal{P})$, such that $II \cup A \vdash Q$ using a strict rule $r = P \leftarrow P, Q_1, \ldots, Q_k$. Then the occurrence of $P$ in the antecedent can also be proven from $II \setminus \{r\} \cup A$. Thus, there exists a derivation for $Q$ from $II \setminus \{r\} \cup A$ (the same holds the other way round). Therefore, $\langle A, Q \rangle \in \text{Args}(\mathcal{P}')$ iff $\langle A, Q \rangle \in \text{Args}(\mathcal{P} \setminus \{r\})$.

Assume now that $\langle A, P \rangle$ is an argument in $\text{Args}(\mathcal{P})$, such that $II \cup A \vdash P$ using a defeasible rule $r = P \leftarrow P, S_1, \ldots, S_k$. Let $A' = A \setminus \{r\}$. Clearly, $II \cup A' \vdash P$. But then $\langle A, P \rangle$ is not an argument, since it is not minimal (contradiction). Therefore, no defeasible rule $P \leftarrow P, S_1, \ldots, S_k$ can be used in building an argument. Therefore, $\langle A, P \rangle \in \text{Args}(\mathcal{P})$ iff $\langle A, P \rangle \in \text{Args}(\mathcal{P} \setminus \{r\})$. $\square$

It must be remarked that defeasible information in a given argument is represented through the defeasible rules used in its construction. This explains why we have to restrict ourselves to strict rules when considering $\text{SUB}\textsubscript{not}$ and $\text{UNFOLD}\textsubscript{not}$. Performing such transformations on defeasible rules may cause the loss of specificity information present in the antecedent of those rules (i.e., information that distinguishes a defeasible rule as “more informed” than another). A similar situation will arise with respect to $\text{SUB}\textsubscript{not}$ and $\text{UNFOLD}\textsubscript{not}$, as presented in Section 4.2.

4.2. Transformation properties in DeLP\textsubscript{not}

DeLP\textsubscript{not} is the subclass of programs in DeLP which contain only default negation not, but no strict negation “~”. This class can also be seen as NLP with the addition of defeasible rules. In such a setting there is no strict negation “~”, and therefore no contradictory literals $P$ and $\sim P$ can appear. The attack relationship among arguments is defined in terms of default literals: an argument $\langle A, Q_1 \rangle$ accounts for a counterargument for an argument $\langle B, Q_2 \rangle$ if not $Q_1$ is used as an assumption in the defeasible derivation of $Q_2$ from $II \cup B$.

Assumption literals are the only possible points for attack in DeLP\textsubscript{not}. In fact, we now restrict our framework in that we allow in Definition 2.10 only assume $\sim A$ where $A$ is an atom. That is, we do not allow assume $A$ literals (where $A$ is not strictly negated). Thus the set $\mathcal{H}(A, Q)$ denotes in this section the set of assumption literals in $\langle A, Q \rangle$ where all literals are (strictly) negated atoms. The reason is that we want to have as much assume $\sim A$ as is consistently possible: these negated atoms do represent the closed world assumption which is always implicit in such a setting.

An argument involving an assumption assume $\sim A$ will be attacked by any other argument concluding $A$. In order to capture this situation, the notion of a contradictory set of literals has been extended after Definition 2.6 to consider assumption literals.

Strict arguments $\langle \emptyset, R \rangle$ have the special property of defeating any other argument involving an assumption literal, as shown in the following proposition.

Proposition 4.18. Let $\mathcal{P}$ be a DeLP\textsubscript{not} program, and let $\langle A, Q \rangle$ be an argument in $\mathcal{P}$ such that $Q$ follows from $A$ using assume $\sim R$ as an assumption. If $\langle \emptyset, R \rangle$, then $\langle A, Q \rangle$ is not a justification.
Proof. Clearly \( \langle \emptyset, R \rangle \) is a counterargument for \( \langle \mathcal{A}, Q \rangle \), in particular (according to specificity) a defeater. Since \( \langle \emptyset, R \rangle \) has no defeaters (as discussed before), the dialectical tree with root \( \langle \mathcal{A}, Q \rangle \) will have a children node \( \langle \emptyset, R \rangle \), which will turn out to be marked as \( U \) (according to Definition 2.27). Hence \( \langle \mathcal{A}, Q \rangle \) will be marked as \( D \), so that \( \langle \mathcal{A}, Q \rangle \) is not a justification. \( \square \)

The precise semantics for \( \text{DeLP}_{\text{not}} \) depends on the analogue of Definitions 2.14 and 2.18 and the appropriate notion of a dialectical tree. Suitable definitions capture different semantics [18]. But independently of these notions, it can be stated that not \( Q \) will not hold whenever \( Q \) can be ultimately defeated. In particular, not \( Q \) will not hold whenever there is a strict argument for \( Q \). In this respect, \( \text{DeLP}_{\text{not}} \) naturally extends the intended meaning of default negation in traditional logic programming (not \( H \) holds iff \( H \) fails to be finitely proven). This fact also suffices to decide which of the transformation properties are satisfied or to give counterexamples.

Since a \( \text{DeLP}_{\text{not}} \) program does not involve strict negation, many problems considered in Section 4.1 do not arise. New transformations \( \text{RED}^\text{not}_-, \text{RED}^\text{not}_+, \text{SUB}_\text{not}, \text{UNFOLD}_\text{not}, \text{TAUT}_\text{not} \) can be defined, with the same meaning as the ones introduced in Section 4.1 for \( \text{DeLP}_{\text{neg}} \), but referring to \( \text{DeLP}_{\text{not}} \) programs. Similarly, we will use the \( \mathcal{P} \rightarrow_{R^+} \mathcal{P}' \) (resp. \( \mathcal{P} \rightarrow_{R^-} \mathcal{P}', \mathcal{P} \rightarrow_{\text{not}} \mathcal{P}' \), \( \mathcal{P} \rightarrow_{\text{not}, U} \mathcal{P}' \), \( \mathcal{P} \rightarrow_{\text{not}, T} \mathcal{P}' \) ) to denote the \( \text{DeLP}_{\text{not}} \) program \( \mathcal{P}' \) resulting from \( \mathcal{P} \) by application of the transformation \( \text{RED}^\text{not}_+ \) (resp. \( \text{RED}^\text{not}_-, \text{SUB}_\text{not}, \text{UNFOLD}_\text{not}, \text{UNFOLD}_\text{not}, \text{TAUT}_\text{not} \)).

For each transformation, we will show that the resulting transformed program is equivalent to the original one. In the case of \( \text{SUB}_\text{not} \) and \( \text{UNFOLD}_\text{not} \), we restrict ourselves to strict rules, since these transformations do not hold when applied to defeasible rules (as shown in Examples 4.10 and 4.8).

Proposition 4.19. Let \( \mathcal{P} \) be a \( \text{DeLP}_{\text{not}} \) program. Let \( \mathcal{P}' \) be the \( \text{DeLP}_{\text{not}} \) program resulting from \( \mathcal{P} \rightarrow_{R^+} \mathcal{P}' \). Then \( \text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}') \).

Proof. Let \( \mathcal{P} \) be a \( \text{DeLP}_{\text{not}} \) program, such that \( r = P \not\leftarrow Q_1, \ldots, \not\leftarrow Q, \not\leftarrow Q_k \) is a defeasible rule in \( \mathcal{P} \), and there is no rule about \( Q \) in \( \mathcal{P} \). Let \( \mathcal{P}' \) be the \( \text{DeLP}_{\text{not}} \) program resulting from applying \( \mathcal{P} \rightarrow_{R^+} \mathcal{P}' \) to \( \mathcal{P} \) on rule \( r \).

Let \( H \) be an arbitrary literal in \( \mathcal{P} \), such that rule \( r \) is used in building the defeasible derivation of some argument \( \langle \mathcal{A}, S \rangle \), so that \( \text{assume} \sim Q \) is an assumption literal in \( \langle \mathcal{A}, S \rangle \). Since \( \mathcal{P}' = \mathcal{P}\setminus \{r\} \cup \{P \not\leftarrow Q_1, \ldots, Q_k\} \), it is clear that \( S \) has also a defeasible derivation from \( \mathcal{A}\setminus \{r\} \cup \{P \not\leftarrow Q_1, \ldots, Q_k\} \), which is minimal and non-contradictory. Hence, we have the argument \( \langle \mathcal{A}\setminus \{r\} \cup \{P \not\leftarrow Q_1, \ldots, Q_k\}, S \rangle \) in \( \mathcal{P}' \).

Since there is no rule with head \( Q \) in \( \mathcal{P} \), there exists no argument \( \langle C, \mathcal{A} \rangle \) in \( \mathcal{P} \) and hence no counterargument for \( \langle \mathcal{A}, S \rangle \) at \( \text{assume} \sim Q \). Therefore, each defeater for \( \langle \mathcal{A}, S \rangle \) in \( \mathcal{P} \) is also a defeater for \( \langle \mathcal{A}', S \rangle \) in \( \mathcal{P}' \), where \( \mathcal{A}' = \mathcal{A}\setminus \{r\} \cup \{P \not\leftarrow Q_1, \ldots, Q_k\} \). The same line of reasoning applies if \( r \) is a strict rule \( P \not\leftarrow Q_1, \ldots, Q_k \).

Hence each dialectical tree \( \mathcal{T} \) in \( \mathcal{P} \) involving \( \langle \mathcal{A}, S \rangle \) as a node is isomorphic to \( \mathcal{T}' \) in \( \mathcal{P}' \) involving \( \langle \mathcal{A}', S \rangle \) in \( \mathcal{P}' \). From Proposition 4.4 it follows that \( \text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}') \). \( \square \)
Proposition 4.20. Let $\mathcal{P}$ be a DeLP_{not} program. Let $\mathcal{P}'$ be the DeLP_{not} program resulting from $\mathcal{P} \rightarrow_{K_{\notnot}} \mathcal{P'}$. Then $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$.

Proof. Let $\mathcal{P} = (\Pi, \Lambda)$ be a DeLP_{not} program. Let $r = P \leftarrow Q_1, \ldots, Q_n$ be a strict rule, and assume $Q \not\in \mathcal{P}$. Assume $r$ is used in a defeasible derivation for building an argument $\langle \mathcal{A}, H \rangle$. Clearly, $\Pi \cup \mathcal{A} \vdash Q$ and $\Pi \cup \mathcal{A} \vdash \neg Q$. But this violates condition 2 in Definition 2.10 (contradiction). Therefore, each argument $\langle \mathcal{A}, H \rangle$ in $\mathcal{P}$ is also an argument in $\mathcal{P} \setminus \{r\}$. Hence $\text{Args}(\mathcal{P}) = \text{Args}(\mathcal{P}')$, and therefore $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$. □

Proposition 4.21. Let $\mathcal{P}$ be a DeLP_{not} program. Let $\mathcal{P}'$ be the DeLP_{not} program resulting from $\mathcal{P} \rightarrow_{\notnot} \mathcal{P}'$. Then $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$.

Proof. Let $\mathcal{P}$ be a DeLP_{not} program, and let $r = P \leftarrow B_1$ be a non-minimal strict rule in $\mathcal{P}$ (i.e., there exists a rule $r' = P \leftarrow B_2$ such that $B_2 \subseteq B_1 \setminus \{r\}$). We consider $B_1 = B_1^+ \cup \text{not } B_1^-$, distinguishing the set $B_1^+$ of positive literals from the set not $B_1^-$ (literals preceded by not). If $B_2 \subseteq B_1$, then two situations are to be considered: either $B_2^+ \subseteq B_1^+$, or $B_2^- \subseteq B_1^-$. (1) Suppose $B_2^+ \subseteq B_1^+$. Then $\text{Args}(\mathcal{P}) = \text{Args}(\mathcal{P} \setminus \{r\})$, following the same line of reasoning as in Proposition 4.9. (2) Suppose that $B_2^- \subseteq B_1^-$, $B_2^+ = B_1^+$. Suppose there exists an argument $\langle \mathcal{A}, H \rangle$ such that the strict rule $r = P \leftarrow B_1$ is used in the defeasible derivation of $H$. Clearly, there is an assumption literal $\text{assume } \neg Q$ in $\mathcal{A}$ for each not $Q$ in $B_1^-$. Let $\mathcal{H}_1$ be the set of assumption literals in $\mathcal{A}$. It follows that $\mathcal{A} \setminus \mathcal{H}_2$ also provides a defeasible derivation for $H$ using $r'$ instead, where $\mathcal{H}_2$ is the set of assumption literals in $r'$, such that $\mathcal{H}_2 \subseteq \mathcal{H}_1$. But then the defeasible derivation of $H$ using $r$ violates condition 3 in Definition 2.10. Therefore, no argument using $r$ can be built in $\mathcal{P}$, so that $\text{Args}(\mathcal{P}) = \text{Args}(\mathcal{P} \setminus \{r\})$. From this analysis it follows that $\mathcal{P} \rightarrow_{\notnot} \mathcal{P}'$ is such that $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$. □

Proposition 4.22. Let $\mathcal{P}$ be a DeLP_{not} program. Let $\mathcal{P}'$ be the DeLP_{not} program resulting from $\mathcal{P} \rightarrow_{T_{\notnot}} \mathcal{P}'$. Then $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$.

Proof. We will consider only the case in which literals preceded by not are present in a rule of the form $r = P \leftarrow P, Q_1, \ldots, Q_k$. Otherwise the proof follows the same line of reasoning as in Proposition 4.17. (1) Suppose there exists an argument $\langle \mathcal{A}, H \rangle$ in $\mathcal{P}$ such that $\Pi \cup \mathcal{A} \vdash H$ using a strict rule $r = P \leftarrow P, Q_1, \ldots, Q_k$. Then the occurrence of $P$ in the antecedent of $r$ can also be proven from $\Pi \setminus \{r\} \cup \mathcal{A}'$, where $\mathcal{A}' = \mathcal{A} \setminus \{\text{assume } \neg Q\}$. But then $\langle \mathcal{A}, H \rangle$ is not an argument, since it violates condition 3 in Definition 2.10. Therefore, no rule $r = P \leftarrow P, Q_1, \ldots, Q_k$ can be used in an argument in $\mathcal{P}$. Hence $\text{Args}(\mathcal{P}) = \text{Args}(\mathcal{P}')$, with $\mathcal{P}' = \mathcal{P} \setminus \{r\}$ so that $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$. (2) Suppose there exists an argument $\langle \mathcal{A}, H \rangle$ in $\mathcal{P}$ such that $\Pi \cup \mathcal{A} \vdash H$ using a defeasible rule $r = P \leftarrow P, Q_1, \ldots, not Q, \ldots, Q_k$. The same line of reasoning
as above applies, with $\mathcal{A}' = \mathcal{A} \setminus \{r, \text{assume } \sim Q\}$. Therefore, $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$. □

We present now a property of immediate subarguments in $\text{DeLP}_{\text{not}}$, similar to the one shown in Proposition 4.14. Then we will show that the transformation $\not\cup_{\text{not}}$ preserves semantics when applied to a given $\text{DeLP}_{\text{not}}$ program.

**Proposition 4.23.** Let $\langle \mathcal{A}, H \rangle$ be an argument in $\text{DeLP}_{\text{not}}$, such that the last rule used in the derivation is the strict rule $H \leftarrow P_1, \ldots, P_k, \text{not } L_1, \ldots, \text{not } L_j$, distinguishing literals from assumption literals. Then all immediate subarguments $\langle \mathcal{A}_i, P_1 \rangle, \ldots, \langle \mathcal{A}_k, P_k \rangle$ are such that $\mathcal{A}_i = \mathcal{A} \setminus \cup_{j=1}^i \{\text{assume } \sim L_i\}$, $\forall i = 1, \ldots, k$.

**Proof.** Follows from the same line of reasoning in Proposition 4.14 when considering strict rules with assumption literals. □

**Proposition 4.24.** Let $\mathcal{P}$ be a $\text{DeLP}_{\text{not}}$ program. Let $\mathcal{P}'$ be the $\text{DeLP}_{\text{not}}$ program resulting from $\mathcal{P} \not\cup_{\text{not}} \mathcal{P}'$ wrt a strict rule $r$ in $\mathcal{P}$. Then $\text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}')$.

**Proof.** Let $\mathcal{P}_1 = (\Pi, \Delta)$ be a $\text{DeLP}_{\text{not}}$ program, and let $\langle \mathcal{A}, Q \rangle$ be an argument in $\mathcal{P}_1$, such that (i) a strict rule $r = H \leftarrow \mathcal{B}$ is used in the defeasible derivation of $Q$ from $\Pi \cup \mathcal{A}$, and (ii) $r$ is $\text{UNFOLD}_{\text{not}}$-related to other rule $r_i$. If this is not the case, then clearly $\text{Args}(\mathcal{P}_1) = \text{Args}(\mathcal{P}_2)$, and the proposition holds trivially. We can also assume that $\emptyset \subseteq \mathcal{B} \subseteq \mathcal{A}$, i.e., there is at least one literal preceded by $\text{not }$ in $\mathcal{B}$; otherwise the proposition follows directly from Proposition 4.15.

Since rule $r$ was applied in the defeasible derivation of $Q$ from $\Pi \cup \mathcal{A}$, there exists an argument $\langle \mathcal{F}, H \rangle$ which is a subargument of $\langle \mathcal{A}, Q \rangle$, such that the last rule used in the defeasible derivation of $\langle \mathcal{F}, H \rangle$ is $r$.

The strict rule $r$ can be written as

$$r = H \leftarrow B, L_1, \ldots, L_k, \text{not } M_1, \ldots, \text{not } M_j$$

(4)

distinguishing positive literals from literals preceded by $\text{not }$. Let $\mathcal{F}_1 = \mathcal{F} \setminus \cup_{i=1}^j \{\text{assume } \sim M_i\}$. From Proposition 4.23, we get that $\langle \mathcal{F}_1, B \rangle, \langle \mathcal{F}_1, L_1 \rangle, \ldots, \langle \mathcal{F}_1, L_k \rangle$ are immediate subarguments of $\langle \mathcal{F}, H \rangle$. Hence we get that

$$\mathcal{H}(\langle \mathcal{F}, H \rangle) = \mathcal{H}(\langle \mathcal{F}_1, B \rangle) \cup \bigcup_{i=1}^k \mathcal{H}(\langle \mathcal{F}_1, L_i \rangle) \cup \bigcup_{i=1}^j \{\text{assume } \sim M_i\}$$

(5)

Consider $r_i = B \leftarrow \mathcal{B}$, which is the last rule used in the defeasible derivation of $\langle \mathcal{F}_1, B \rangle$, such that $r_i$ is $\text{UNFOLD}_{\text{not}}$-related to $r_i$. Since $r_i$ is an arbitrary strict rule, it will have the form

$$r_i = B \leftarrow P_1, \ldots, P_m, \text{not } R_1, \ldots, \text{not } R_p$$

(6)

Let $\mathcal{F}_2 = \mathcal{F}_1 \cup \cup_{i=1}^p \{\text{assume } \sim R_i\}$. It follows that

$$\mathcal{H}(\langle \mathcal{F}_1, B \rangle) = \bigcup_{i=1}^m \mathcal{H}(\langle \mathcal{F}_2, P_i \rangle) \cup \bigcup_{j=1}^p \{\text{assume } \sim R_j\}$$

(7)
Replacing (7) in (5), we get

\[ H(\mathcal{F}, H) = \bigcup_{i=1}^{m} H(\mathcal{F}, P_i) \cup \bigcup_{j=1}^{p} \{ \text{assume} \sim R_j \} \cup \bigcup_{i=1}^{k} H(\mathcal{F}, L_i) \]

\[ \cup \bigcup_{j=1}^{p} \{ \text{assume} \sim M_i \} \quad (8) \]

Thus, argument \( \langle \mathcal{F}, H \rangle \) in \( \mathcal{P}_1 \) is such that \( \mathcal{F} = \mathcal{R}_S \cup H(\mathcal{F}, H) \), where \( H(\mathcal{F}, H) \) is defined as in (8). Assume we apply \( \sim_{\text{UNFOLD}} \) to \( \mathcal{P}_1 \), where the rule \( r \) is \( \text{UNFOLD}_{\text{not}} \)-related to \( r_i \), resulting in a new \( \text{DeLP}_{\text{not}} \) program \( \mathcal{P}_2 \). From the definition of \( \text{UNFOLD}_{\text{not}}^\prime \), we have:

\[ \mathcal{P}_2 = \mathcal{P}_1 \setminus \{ H \leftarrow B \} \cup \{ H \leftarrow (B \setminus \{ B \}) \cup B' \} | r_i \leftarrow B' \in \mathcal{P}_1 \}. \]

Consider the original rule \( r \) in (4), and the \( \text{UNFOLD}_{\text{not}} \)-related rule \( r_i \) in (6). Let \( \mathcal{P}_2 \) be the \( \text{DeLP}_{\text{not}} \) program resulting from applying the UNFOLD transformation to \( r \) with respect to \( r_i \). In this case we get

\[ H \leftarrow \{ B, L_1, \ldots, L_k, \text{not} M_1, \ldots, \text{not} M_j \} \setminus \{ B \} \cup \{ P_1, \ldots, P_m, \text{not} R_1, \ldots, \text{not} R_p \} \]

or equivalently

\[ H \leftarrow L_1, \ldots, L_k, \text{not} M_1, \ldots, \text{not} M_j, \text{not} R_1, \ldots, \text{not} R_p \quad (9) \]

Let \( \mathcal{F}' = \mathcal{F} \setminus \{ \cup_{i=1}^{j} \{ \text{assume} \sim M_i \} \cup \cup_{j=1}^{p} \{ \text{assume} \sim R_j \} \} \). From Proposition 4.23, it follows that \( \langle \mathcal{F}', L_i \rangle, i = 1, \ldots, k \) and \( \langle \mathcal{F}', P_i \rangle, i = 1, \ldots, m \) are arguments in \( \mathcal{P}_2 \). In particular, we have

\[ H(\mathcal{F}', H) = \bigcup_{i=1}^{k} H(\mathcal{F}', L_i) \cup \bigcup_{i=1}^{m} H(\mathcal{F}', P_i) \]

\[ \cup \bigcup_{j=1}^{p} \{ \text{assume} \sim R_j \} \cup \bigcup_{i=1}^{j} \{ \text{assume} \sim M_i \}. \quad (10) \]

Hence \( \mathcal{R}_S \cup H(\mathcal{F}', H) \) is an argument for \( H \) in \( \mathcal{P}_2 \), since every defeasible rule in \( \mathcal{P}_1 \) is also a defeasible rule in \( \mathcal{P}_2 \). But from (8) and (10) it follows that \( H(\mathcal{F}', H) = \mathcal{H}(\mathcal{F}, H) \), and the set \( \mathcal{F}' = \mathcal{F} \). Hence, \( \langle \mathcal{F}, H \rangle \) is an argument in both \( \mathcal{P}_1 \) and \( \mathcal{P}_2 \).

Therefore, we can conclude that for any argument \( \langle \mathcal{F}, H \rangle \) in \( \mathcal{P}_1 \) such that one of the strict rules \( r \) used in its defeasible derivation is \( \text{UNFOLD}_{\text{not}} \)-related to another rule \( r_i \), it follows that \( \langle \mathcal{F}, H \rangle \) is also an argument in \( \mathcal{P}_2 \). Note that no new argument other than \( \langle \mathcal{F}, H \rangle \) is generated in \( \mathcal{P}_2 \), since the subarguments of \( R(\mathcal{F}, H) \) in \( \mathcal{P}_1 \) and \( \langle \mathcal{F}, H \rangle \) in \( \mathcal{P}_2 \) are the same. Hence \( \text{Args}(\mathcal{P}_1) = \text{Args}(\mathcal{P}_2) \), and therefore \( \text{SEM}_{\text{DeLP}}(\mathcal{P}_1) = \text{SEM}_{\text{DeLP}}(\mathcal{P}_2) \). \( \square \)

**Corollary 4.25.** Let \( \mathcal{P} \) be a \( \text{DeLP}_{\text{not}} \) program. Let \( \mathcal{P}' \) be the \( \text{DeLP}_{\text{not}} \) program resulting from \( \mathcal{P} \sim_{\text{UNFOLD}} \mathcal{P}' \) wrt a strict rule \( r \) in \( \mathcal{P} \). Then \( \text{SEM}_{\text{DeLP}}(\mathcal{P}) = \text{SEM}_{\text{DeLP}}(\mathcal{P}') \).

**Proof.** Follows directly from Proposition 4.13 by repeated application of \( \sim_{\text{UNFOLD}} \), for each \( r_i \) which is \( \text{UNFOLD}_{\text{not}} \)-related with \( r \). \( \square \)
4.3. Relating NLP and DeLP\textsubscript{not} under WFS and Stable

A natural question is how well-founded semantics WFS relates to DeLP\textsubscript{not}. The answer is very simple because of our results that the transformation properties are semantics preserving and the fact that programs in normal form have an obvious semantics.

**Theorem 4.26 (DeLP\textsubscript{not} extends WFS).** Let \( \mathcal{P} \) be a program in NLP. We can look at \( \mathcal{P} \) as a theory in DeLP\textsubscript{not}. Then all atoms \( A \) and default atoms \( \neg A \) that are true in WFS(\( \mathcal{P} \)) are also contained in \( \text{SEM}_{\text{DeLP}\textsubscript{not}}(\mathcal{P}) \).

**Proof.** As all the transformation properties hold, we can transform \( \mathcal{P} \) into a normal form where all rules only have negative body literals (or are empty). Note that we do not have defeasible rules in this context, because we consider simple nlp\textsubscript{s} only.

- The atoms true in WFS(\( \mathcal{P} \)) are, by Theorem 3.4, exactly those \( A \) where there is a rule of the form “\( A \leftarrow \)”. But those atoms are certainly justified in \( \text{SEM}_{\text{DeLP}\textsubscript{not}}(\mathcal{P}) \).
- All default literals \( \neg A \) that are true in WFS(\( \mathcal{P} \)) are, by Theorem 3.4, exactly those \( A \) where there is no rule with head \( A \). But then assume \( \sim A \) can be assumed as it cannot lead to any contradiction.

**Example 4.27.** Consider the normal logic programs

\[
\begin{align*}
\mathcal{P}_1 &= \{ (a \leftarrow b), (b \leftarrow a), (c \leftarrow \neg a, \neg b) \} \\
\mathcal{P}_2 &= \{ (a \leftarrow \neg b), (b \leftarrow a) \} \\
\mathcal{P}_3 &= \{ (a \leftarrow \neg b), (b \leftarrow \neg a), (c \leftarrow a), (c \leftarrow b) \} \\
\mathcal{P}_4 &= \{ (a \leftarrow b, \neg d), (b \leftarrow a, \neg d), (d \leftarrow \neg d), (c \leftarrow \neg a, \neg b) \} \\
\mathcal{P}_5 &= \{ (a \leftarrow \neg b), (b \leftarrow \neg a), (a \leftarrow \neg a) \} \\
\mathcal{P}_6 &= \{ (a \leftarrow \neg b), (b \leftarrow \neg c), (c \leftarrow \neg d), (d \leftarrow \neg e), (e \leftarrow \neg d) \}
\end{align*}
\]

We analyze the above NLP programs as DeLP\textsubscript{not} programs.

- WFS in \( \mathcal{P}_1 \) is \{\( \neg a, \neg b, c \}\}. The only argument that can be constructed from \( \mathcal{P}_1 \) as a DeLP\textsubscript{not} program is the one which justifies \( c \). Without the last rule (\( c \leftarrow \neg a, \neg b \)) no arguments for positive atoms can be constructed.
- WFS in \( \mathcal{P}_2 \) is empty. Under DeLP\textsubscript{not}, no argument can be built, since the only possible set \{\assumption \sim b\} leads to contradiction.
- WFS in \( \mathcal{P}_3 \) is empty. In DeLP\textsubscript{not}, two sets of assumptions are possible for building arguments: \( \mathcal{A}_1 = \{ \assumption \sim a \} \) and \( \mathcal{A}_2 = \{ \assumption \sim b \} \). We can build the arguments \( \langle \mathcal{A}_1, b \rangle, \langle \mathcal{A}_2, a \rangle, \langle \mathcal{A}_1, c \rangle, \langle \mathcal{A}_2, c \rangle \). Any one of these arguments has a blocking defeater (\( \mathcal{A}_1 \) blocks \( \mathcal{A}_2 \) and vice versa). From Definition 2.28 it follows that no argument is justified.
- WFS in \( \mathcal{P}_4 \) is \{\( \neg a, \neg b, c \}\}. The only argument that can be constructed from \( \mathcal{P}_4 \) as a DeLP\textsubscript{not} program is the one which justifies \( c \). However, without the last
rule $c \leftarrow \text{not } a, \text{not } b$ no argument can be built in $\mathcal{P}_4$ under $\text{DeLP}_{\text{not}}$ (there is no defeasible sequence for $a$ nor for $b$).

- WFS in $\mathcal{P}_5$ is empty. There are no stable models for this program. But in $\text{SEM}_{\text{DeLP}_{\text{not}}}$ the argument $\{\text{assume } \sim b\}$ is a justification for $a$. This is because $\langle \{\text{assume } \sim b\}, a \rangle$ cannot be defeated (the only way to do this would be to find an argument involving the assumption $\text{not } a$, but this would lead to a contradiction).

- WFS in $\mathcal{P}_6$ is empty, but there exist two stable models: $\{b, d\}$ and $\{a, c, e\}$. In $\text{SEM}_{\text{DeLP}_{\text{not}}}$ the argument $\{\text{assume } \sim b\}$ is not a justification for $a$, because the second rule acts as a (blocking) defeater for this argument. Note that it is irrelevant that the third rule is also a blocking defeater of the second: this is excluded in condition 3 of Definition 2.21. There are no justifications at all for this program. Consequently, $\text{SEM}_{\text{DeLP}_{\text{not}}}$ is empty.

The last two programs show that $\text{SEM}_{\text{DeLP}_{\text{not}}}$ is strictly stronger than WFS. In the following, we will show in addition that Stable is stronger than $\text{SEM}_{\text{DeLP}_{\text{not}}}$ (Theorem 4.28). This relation between the two semantics again is strict. In order to see this, consider $\mathcal{P}$ simply defined as $\mathcal{P}_3$ above: in both stable models of $\mathcal{P}_3$ (namely $\{\text{not } a, b, c\}$ and $\{\text{not } b, a, c\}$) the atom $c$ holds. However, there is no justification for $c$ in $\mathcal{P}$ as a $\text{DeLP}_{\text{not}}$ theory (because the arguments $\langle \{\text{assume } \sim a\}, c \rangle$ and $\langle \{\text{assume } \sim b\}, c \rangle$ block each other).

**Theorem 4.28.** Let $\mathcal{P}$ be a nlp. Then, $\text{SEM}_{\text{DeLP}}(\mathcal{P}) \subseteq \text{Stable}(\mathcal{P})$. Here, Stable($\mathcal{P}$) denotes the set of literals that hold in all stable models of $\mathcal{P}$.

**Proof.** If there is no stable model for $\mathcal{P}$ at all, then Stable($\mathcal{P}$) contains all literals. Hence, trivially the conjecture holds in this case. Thus, in the sequel, we will assume that $\mathcal{P}$ has at least one stable model. Furthermore, since both for Stable and $\text{SEM}_{\text{DeLP}_{\text{not}}}$ all transformation rules (in Fig. 3) hold, we can restrict our attention to nlps that are in normal form. Note that there cannot be any counterarguments at all, because we have only blocking defeaters for nlps which defeat any argument by condition 3 in Definition 2.23.

Let us now make a case distinction:

1. An atom $a$ occurring as a fact in $\mathcal{P}$ is clearly justified wrt the $\text{DeLP}$ semantics and must also be contained in any stable model of $\mathcal{P}$ (because stable models are
models of the underlying program). Hence, the theorem holds for such positive literals \( a \).

(2) For atoms \( a \) not occurring in any head in the program \( \mathcal{P} \), there cannot be any argument. Hence, these atoms belong to \( \mathcal{P}^{\text{projected}} \). Clearly, not \( a \) must be in any stable model of \( \mathcal{P} \) (because stable models are grounded) which completes this part of the proof.

(3) We do not have to consider atoms \( a \), for which there are arguments, all of which have blocking defeaters, because they belong to \( \mathcal{P}^{\text{undefined}} \) (see Definition 2.31). Therefore, neither \( a \) nor not \( a \) is contained in \( \text{SEM}_{\text{DeLP}}(\mathcal{P}) \).

(4) The only case remaining are atoms \( a \), that occur in rules with head \( a \) and a negative body, say \( a \leftarrow \text{not } b_1, \text{not } b_i \), and there is an argument for \( a \) which is not defeated. Note that all the \( b_i \) do also occur as heads of other rules, say \( b_i \leftarrow \text{not } c_{i_1}, \text{not } b_{i_0} \), so they are potential (blocking) defeaters. The only reason that these rules are not defeaters (and thus there is an argument for \( a \)) is that the sets \{not \( c_{i_1}, \text{not } b_{i_0} \} \) are all inconsistent with the program: i.e., condition 2 of Definition 2.10\(^4\) is violated. That means that in each stable model (if they exist) all the \( b_i \) must be false (because stable models are two-valued and grounded) and thus \( a \) must be true. □

Putting the results of the Theorems 4.26 and 4.28 together, we get that, for nlps, the \( \text{DeLP} \) semantics lies between WFS and Stable. This is summarized in the following corollary.

**Corollary 4.29.** Let \( \mathcal{P} \) be a nlp. Then:

\[
\text{WFS}(\mathcal{P}) \subsetneq \text{SEM}_{\text{DeLP}}(\mathcal{P}) \subsetneq \text{Stable}(\mathcal{P})
\]

### 4.4. Relating NLP and DeLP: Summary

Fig. 3 summarizes the behavior of \( \text{NLP} \), \( \text{DeLP}_{\text{neg}} \) and \( \text{DeLP}_{\text{not}} \) under the different transformation rules presented before. From that table, we can identify some relevant features:

- An argumentation-based semantics has been given to \( \text{NLP} \) using an abstract argumentation framework by Kakas and Toni [22]. From Section 4.2 it is clear that \( \text{DeLP} \) is a proper extension of \( \text{NLP} \), since there are transformation properties in \( \text{NLP} \) which do not hold in \( \text{DeLP} \). This is basically due to the knowledge representation capabilities provided by defeasible rules.

- Some properties of \( \text{NLP} \) under well-founded semantics are also present in \( \text{DeLP} \) (such as \text{TAUT} and \text{RED}^-). It is worth noticing that \text{RED}^- holds in \( \text{NLP} \) because of a “consistency constraint” (it cannot be the case that both not \( a \). The same is achieved in \( \text{DeLP} \) by demanding non-contradiction when constructing arguments.

- Other transformation properties only hold for strict rules (e.g., \text{SUB}), sometimes with extra requirements (e.g., \text{UNFOLD}). This shows that defeasible rules express

\(^4\) One example for this is program \( \mathcal{P}_5 \) above.
a link between literals that cannot be easily “simplified” in terms of a transformation rule, and a more complex analysis (e.g., computing defeat) is required.

- Some properties (e.g., RED⁺) do not hold wrt strict negation, but do hold wrt default negation. In the first case, the reason is that negated literals are treated as new predicate names (and succeed as subgoals iff they can be proven from the program). In the second case, default negation behaves much like its counterpart in NLP. As in NLP, the absence of rules with head H is enough for concluding that H cannot be proven, and therefore not justified.

5. Related work and conclusion

5.1. Related work

In recent work [22], an abstract argumentation framework has been used as a basis for defining an unifying proof theory for various argumentation semantics of logic programming. In that framework, well-founded semantics for NLP is computed by using an argument-based approach, which has many similarities with DeLP [7].

Many semantics for extended logic programs view default negation and symmetric negation as unrelated. To overcome this situation a semantics well-founded semantics with explicit negation (WFSX) for extended logic programs has been defined by Alferes et al. [1]. WFSX embeds a “coherence principle” providing the natural missing link between both negations: if ~L holds then not L should hold too (similarly, if L then not ~L). In DeLP this “coherence principle” also holds [18].

Finally, it must be remarked the original Simari–Loui formulation [31] contains a fixed-point definition that characterizes all justified beliefs. A similar approach was used later by Prakken and Sartor [27] in an extended logic programming setting, getting a revised version of well-founded semantics as defined by Dung [15]. These analogies highlight the link between well-founded semantics and skeptical argumentative frameworks.

5.2. Conclusion

We have related in this article the logical framework DeLP to classical logic programming semantics, particularly well-founded semantics for NLP. The link between both semantics was established by looking for analogies and differences in the results of applying transformation rules on logic programs.

The differences between NLP and DeLP are to be found in the expressive power of DeLP for encoding knowledge in comparison with NLP. Defeasible rules allow the formalization of criteria for defeat among arguments which cannot be easily “compressed” by applying transformation rules, as explained in Section 4.4. Strict negation in DeLP is also a feature which extends the representation capabilities of NLP. However, as already discussed, the same principle which guides the application of the transformation rule RED⁻ in NLP can be used for detecting rules that cannot be used for constructing arguments.
It is worth noticing that the original motivation for DeLP was to find an argumentative formulation for defeasible theories in order to resolve potential inconsistencies. This was at the end of the 1980s. In the meantime, the area of semantics for logic programs underwent a solid foundational phase and today several possible semantics together with their properties are well known. We think that these results can be applied to gain a better understanding of argumentation-based frameworks.

Acknowledgements

We thank two anonymous referees for a number of suggestions that helped to improve this article. Special thanks to our dear colleague Pablo Fillottrani for a lot of useful comments and remarks.

References


