Defeasibility in Answer Set Programs with Defaults and Argumentation Rules

Hui Wan, Michael Kifer, Benjamin Grosof

Abstract. Defeasible reasoning has been studied extensively in the last two decades and many different and dissimilar approaches are currently on the table. This multitude of ideas has made the field hard to navigate and the different techniques hard to compare. Our earlier work on Logic Programming with Defaults and Argumentation Theories (LPDA) introduced a degree of unification into the approaches that rely on the well-founded semantics. The present work takes this idea further and introduces ASPDA (Answer Set Programs via Argumentation Rules) — a unifying framework for defeasibility of disjunctive logic programs under the Answer Set Programming (ASP). Since the well-founded and the answer set semantics underlie almost all existing approaches to defeasible reasoning in Logic Programming, LPDA and ASPDA together can closely approximate most of those approaches. In addition to ASPDA, we obtained a number of interesting and non-trivial results. First, we show that ASPDA is reducible to ordinary ASP programs. Second, we study reducibility of ASPDA to the non-disjunctive case and show that head-cycle-free ASPDA programs reduce to the non-disjunctive case — similarly to head-cycle-free ASP programs, but through a more complex transformation. We also shed light on the relationship between ASPDA and some of the earlier theories such as Defeasible Logic and LPDA.

Keywords: Logic Programming, Defeasible Reasoning, Argumentation Theory, Argumentation Rules, Answer Sets, Stable Model

1. Introduction

Defeasible reasoning is a form of non-monotonic reasoning where logical axioms are true “by default” but their truth status may be undercut or even negated by other, conflicting axioms. This type of reasoning has been an important application of logic programming. It was applied to model policies, regulations, and law; actions, change, and process causality; Web services; and aspects of inductive/scientific learning. However, there is a bewildering multitude of dissimilar and incompatible approaches to defeasibility based on a wide variety of intuitions and techniques. The difficulties in relating and comparing the different approaches have been discussed in among others. Combining the various theories of defeasible reasoning with other advances in logic-based knowledge representation, such as HiLog and F-logic, has also been a problem.

Our earlier work addressed some of these issues by introducing a general framework for defeasible reasoning, called LPDA, which abstracts the intuitions about defeasibility into what we call argumentation theories. In LPDA, an argumentation theory is a set of logic axioms that express the arguments for or against defeating various rules in the knowledge base. These arguments often depend on the particular appli-
cation domain and user intent. An argumentation theory should be viewed not as part of a knowledge base but rather of its semantics. This approach enables a uniform syntax and semantics for a wide variety of defeasible theories, which could be used in harmony and simultaneously in the same knowledge base. LPDA, as defined in [40], was developed on the basis of the well-founded models [13] and was able to unify a number of approaches to defeasible reasoning that are based on the well-founded semantics. However, a large number of works on defeasible reasoning are based on the stable model semantics. Furthermore, a general defeasible reasoning in the presence of disjunctive information appears to require even more general semantics, the answer set semantics [19].

The present work takes the idea of LPDA further and introduces ASPDA—an analogous framework for defeasibility of disjunctive logic rules through argumentation rules [1] based on Answer Set Programming (ASP). In this way, LPDA and ASPDA together unify and extend most of the existing theories of defeasible reasoning in Logic Programming. Extension of the semantics of LPDA to ASP with head-disjunctions turned out to be elegant but not straightforward. The relationship between ASPDA and the regular ASP also proved to be non-obvious. First, we show that ASPDA can be expressed by regular ASP programs. A polynomial reduction has been recently given in [16]. Then we study the class of head-cycle-free programs with disjunctive heads and show that a related notion exists for ASPDA. By analogy with the classical case, such programs can be reduced to non-disjunctive programs under the defeasible stable model semantics, although the transformation is more complicated than in the case of the regular ASP. The blowup in the program size is still linear, however.

To avoid possible confusion, we should mention from the outset that argumentation rules are related to argumentation theories of Dung et al. [14]. We briefly discuss the relationship in Section 5.

A preliminary report on this work appeared in [41]. Compared to that earlier paper, the present paper develops the main concepts to a fuller extent, provides all proofs, and includes extensive examples of instantiations of ASPDA to illustrate the inner workings of the ASPDA framework.

The rest of this paper is organized as follows. Section 2 illustrates defeasible reasoning under the answer-set semantics using the well-known Turkey Shoot example [32]. Section 3 defines the syntax and semantics of defeasible disjunctive logic programs and presents a number of interesting results about reducibility to the regular logic programming and to the non-disjunctive case. Section 4 gives two examples of argumentation rulesets for ASPDA. One is an adaptation of GCLP [24,40] to ASPDA, a theory that is used in all examples throughout this paper. Another is an argumentation ruleset that closely simulates Defeasible Logic [1]. Sections 5 and 6 discuss related work and conclude the paper.

2. Motivating Example

The example in Figure 1 is adapted from the Texas Turkey Shoot game example in [32]. We use the usual syntax of logic programming with the only difference that rules are tagged with @tag symbols and head-disjunctions are allowed. Variables are prefixed with the symbol “?”. In the scenario described in the example, one of the guns is known to be loaded initially, but it is not known which. The objective is to find a plan to kill the turkey by shooting one or both guns assuming that the shooter can observe the effects of his actions. Let g1 and g2 be the constants representing the guns. Numerals are used in the example to represent time points, and the initial time point is assumed to be 1. For instance, shoot(g1,1) and shoot(g1,2) represent the actions of shooting the gun g1 at time points 1 and 2. In the example, some of the rules have tags, e.g., kpld and sht1, and the predicate #overrides specifies priorities among some of these tagged rules.

We distinguish between the classical-logic-like explicit negation neg and the default negation naf (which in this paper will have the answer-set semantics). Literals L and neg L are assumed to be incompatible and cannot both appear in a consistent model. The predicate #opposes specifies additional contradictions, such as the inability for the turkey to be both dead and alive at the same time.

We can now explain how defeasible reasoning works in the above example. The rule tagged with kpld is a frame persistence axiom stating that a loaded gun stays loaded unless some other action ex-

---

1 Earlier [40] we used the term “argumentation theories” but renamed it to avoid possible confusion with a similar and related term used in Dung et al. [13].
3. Defeasible Reasoning with Argumentation Rules

In this section we introduce the syntax and semantics of disjunctive logic programming where defeasibility is controlled by argumentation rules—sets of axioms (or arguments) that say when and why any particular rule should be considered as defeated and the inference it sanctions as null and void. The main syntactic difference from non-defeasible disjunctive logic programming is that rules now have tags, and the main semantic difference is that these rules can be defeated.

Let \( L \) be a logic language with the usual connectives \( \land \) for conjunction, \( \lor \) for disjunction, and \( :- \) for rule implication; and two negation operators: \( \text{neg} \) for explicit negation and \( \text{naf} \) for default negation. The alphabet of the language consists of: an infinite set of variables, which are shown in the examples as alphanumeric symbols prefixed with the question mark "?"; and a set of constant symbols, which can appear as individuals, function symbols, and predicates. Constants will be shown as alphanumeric symbols that are not prefixed with a "?". We assume that the language includes two special propositional constants, \( t \) and \( f \), which stand for true and false, respectively. We also assume the following order on these propositions: \( f < t \).

We use the standard notion of terms in logic programming. Atomic formulas, also called atoms, can be quite general in form: they can be the usual atoms used in ordinary logic programming; or the higher-order expressions of HiLog \([10]\); or the frames of F-logic \([29]\). A literal has one of the following forms:

- An atomic formula.
- \( \text{neg} A \) and \( \text{naf} A \), where \( A \) is an atomic formula.
- \( \text{naf} \text{neg} A \), where \( A \) is an atomic formula.
- \( \text{naf} \text{naf} L \) and \( \text{neg} \text{neg} L \), where \( L \) is a literal; these are identified with \( L \).

The problem is to infer that by firing one or both guns in succession the shooter can kill the turkey despite the uncertainty in the initial state. Note that due to the disjunctions, other existing logic programming approaches to defeasible reasoning cannot handle the above situation, and this is precisely the motivation for our current work. We will return to this example in Section 4.3 after the necessary theory is developed.
Let $A$ denote an atom. Literals of the form $A$ or $\text{neg } A$ (or literals that reduce to these forms after elimination of double negation) are called naf-free literals; literals that reduce to the form $\text{naf } A$ are called naf-literals.

**Definition 1 (Tagged rule)** A tagged rule in a logic language $L$ is an expression of the form

$$\circ L_1 \lor ... \lor L_k : - \text{Body}$$

where $r$ is a term, called the tag of the rule; $L_1$, ..., $L_k$ ($k \geq 0$) are naf-free literals in $L$, called the head literals of the rule; and Body, called the body of the rule, is a conjunction of literals in $L$. As is common in logic programming, we will often write $A, B$ to represent the conjunction $A \land B$. A rule tag is not a rule identifier: several rules can have the same tag.

A constraint is a special form of rule where $f$ is a single head literal. We will usually omit $f$ in such rules.

A formula is a literal, a Boolean combination of literals using conjunction and disjunction, or a rule.

We will often omit showing rule tags when they are immaterial.

**Definition 2 (Ground terms and rules)** A ground term is a term that contains no variables, a ground literal is a variable-free literal, and a ground rule is a rule that has no variables.

**Definition 3 (ASPDA)** An answer-set program with defaults and argumentation rules (an aspda, for short) in a logic language $L$ is a set of tagged rules in $L$, which can be strict or defeasible. Sets or rules that do not have disjunctions in the head will be called non-disjunctive aspdas. Sometimes we will omit tags when they are immaterial.

Strict rules are used as definite statements about the world. In contrast, defeasible rules represent defeasible defaults whose instances can be “defeated” by other rules. Inferences produced by the defeated rules are “overridden.”

We assume that the distinction between strict and defeasible rules is specified in some way: either syntactically or by means of a predicate. For instance, in Section 4, we use the predicate $\#\text{strict}$ for that purpose.

Aspdas are used in conjunction with argumentation rules, which are sets of rules that define conditions under which some rule instances may be defeated by other rules.

**Definition 4 (Argumentation ruleset)** Let $L$ be a logic language. An argumentation ruleset is a set, $AT$, of strict rules in $L$ of the form $\circ L$. We also assume that the language $L$ includes a binary predicate, $\text{defeated}_{AT}$, which may appear in the heads of some rules in $AT$. When confusion does not arise, we will omit the subscript $AT$.

An aspda $\mathcal{P}$ is said to be compatible with $AT$ if $\text{defeated}_{AT}$ does not appear in the rule heads in $\mathcal{P}$. □

In an argumentation ruleset all rules are strict, by definition. The rules in $AT$ will normally contain other predicates, besides $\text{defeated}_{AT}$, that are used to specify how the rules in $\mathcal{P}$ get defeated. We will see full-fledged examples of argumentation rulesets in Section 4. Note that an argumentation ruleset is also an aspda.

Usually argumentation rules employ the concepts of rule priority and contradictions among facts. Priorities are often specified via predicates, such as $\text{overrides}$, which tell that some rules (or rule instances) have higher priorities than other rules (e.g., $\text{overrides}(\text{rule}_{tag1}, \text{rule}_{tag2})$). Contradictions are commonly expressed via predicates such as $\text{opposes}$, which tell that certain facts cannot be true together (e.g., $\text{opposes}(\text{price}(\text{ball}, 20), \text{price}(\text{ball}, 30))$). The $\text{defeated}$ predicate is then defined in terms of $\text{overrides}, \text{opposes}$, and other predicates. In this paper, we adopt the convention that the predicates defined only by argumentation rules will be prefixed with the $\#$-sign, such as $\#\text{conflict}$, and will be in normal font, except $\#\text{defeated}$, which will be typeset in bold. The predicates used and/or defined both by the argumentation rules and user programs will be prefixed with the $\$-$sign and will be in bold font. Meta-predicates, such as $\text{body}$, will also be set in bold. The predicates defined and used only by user programs

---

2 This is easy to generalize to allow Lloyd-Topor extensions.  
3 This makes it easier to specify priorities and conflicts among groups of rules as opposed to individual rules, as in Figure 4 (look for the tags move and frame).  
4 If $\text{defeated}$ does not occur in the head of any rule then the semantics of aspdas trivially reduce to ordinary logic programming.  
5 In principle, we could allow argumentation rules to be defeasible, but we will not do so in this paper.
will be denoted by alphanumeric symbols and will not be marked in any special way.

In defining the semantics, we assume that the argumentation rules are ground. A grounded version of AT with respect to a compatible aspda $\mathcal{P}$ is obtained by appropriately instantiating the variables and meta-predicates.

Note that the theory developed here permits different subsets of the overall aspda $\mathcal{P}$ to use different sets argumentation rules AT with different $\text{defeated}_{\text{AT}}$ predicates. For instance, our implementation of the argumentation rules for the well-founded semantics in an extended logic programming framework with argumentation rules for the well-founded semantics in an extended version of FLORA-2 supports multiple argumentation rule sets.

### 3.1. Interpretations and Models

**Definition 5 (Herbrand universe and Herbrand base)** Let $\mathcal{P}$ be an aspda and AT an argumentation ruleset over language $\mathcal{L}$.

- The **Herbrand universe** of $\mathcal{P}$, denoted $\mathcal{U}_\mathcal{L}$, is the set of all ground terms built using the constants and function symbols that appear in $\mathcal{L}$. When confusion does not arise, we will simply write $\mathcal{U}$, omitting the language subscript.

- The **Herbrand base** of $\mathcal{P}$, denoted $\mathcal{B}_\mathcal{L}$ (or simply $\mathcal{B}$, when no ambiguity arises), is the set of all ground naf-free literals that can be constructed using the predicates in $\mathcal{L}$.

**Definition 6 (Herbrand interpretation)** A **Herbrand interpretation**, $I$, is a subset of $\mathcal{B}$, i.e., a set of ground naf-free literals. In addition, $I$ must contain $\top$ and must not contain $\bot$.

An interpretation is **inconsistent relative to** an atom $A$ if both $A$ and neg $A$ are in $I$. Otherwise, $I$ is **consistent relative to** $A$. An interpretation is **consistent** if it is consistent relative to every atom and **inconsistent** if it is inconsistent relative to some atom.

Note that all interpretations considered in this paper are Herbrand, so we will often neglect to mention “Herbrandness” explicitly.

**Definition 7 (Truth valuation)** Let $I$ be a Herbrand interpretation, $L$ a ground naf-free literal, and let $F$, $G$ be ground formulas. We define truth valuations that map formulas to $\{t, f\}$ as follows:

- $I(L) = t$ if $L \in I$, $I(L) = f$ otherwise.
- $I(\text{\texttt{naf}} L) = \sim I(L)$, where $\sim t = f$ and $\sim f = t$.
- $I(F \land G) = \min(I(F), I(G))$. Recall that $f < t$.
- $I(F \lor G) = \max(I(F), I(G))$.
- For a strict rule $\preceq_r F : - G$, we define $I(F : - G) = t$ if and only if $I(F) \geq I(G)$.

Intuitively, a strict rule is true if its head is “more true” than the body, i.e., either the head is true or the body is false.

- For a defeasible rule $\preceq_r L_1 \lor \ldots \lor L_k : - G$, we define $I(\preceq_r L_1 \lor \ldots \lor L_k : - G) = t$ if and only if $I(L_1 \lor \ldots \lor L_k) \geq \min(I(G), V)$ where $V = \max_{1 \leq i \leq k} I(\text{\texttt{naf}} \text{\texttt{defeated}}(r, L_i))$.

That is, a defeasible rule $\preceq_r$ is true if either (i) it is “defeated,” i.e., $\text{\texttt{defeated}}(r, L_i)$ holds for all $L_i$; or (ii) its body is false; or (iii) if its head is true.

**Definition 8 (Model of formula and rule)** If $F$ is a ground formula, $I$ an interpretation, and $I(F) = t$, then we write $I \models F$ and say that $I$ is a model of $F$ or that $F$ is satisfied in $I$.

If $R$ is a ground rule $\preceq_r L_1 \lor \ldots \lor L_k : - G$, an interpretation, and $I(R) = t$, then we write $I \models R$ and say that $I$ is a model of $R$ or that $R$ is satisfied in $I$.

We write $I \models P$ if $I \models R$ for every $R \in \mathcal{P}$.

**Definition 9 (Model of aspda w.r.t. argumentation theory)** Given an aspda $\mathcal{P}$, an argumentation ruleset AT, and an interpretation $M$, we say that $M$ is a model of $\mathcal{P}$ with respect to the argumentation ruleset AT (or a model of $(\mathcal{P}, \text{AT})$, for short), written as $M \models (\mathcal{P}, \text{AT})$, if $M \models \mathcal{P}$ and $M \models \text{AT}$.

**Definition 10 (Minimal model)** A minimal model of $(\mathcal{P}, \text{AT})$ is a model $M$ of $(\mathcal{P}, \text{AT})$ such that no proper subset of $M$ is a model of $(\mathcal{P}, \text{AT})$.

### 3.2. Stable Model and Answer-set Semantics

In this section, we extend the stable model semantics [20] and the answer-set semantics [19] to ASPDA. We start with non-disjunctive aspda and stable models.

**Definition 11 (ASPDA quotient, non-disjunctive case)** Let $Q$ be a non-disjunctive aspda, and let $J$ be a Herbrand interpretation for $Q$. The **ASPDA quotient $Q$ by $J$**, written as $\frac{Q}{J}$, is defined by the following sequence of steps:

6 In regular ASP theory, the term reduct is normally used. We later use the term reduction in a different sense, so quotient is used here to avoid confusion.
1. Delete every rule \( R \in Q \) such that there is a \text{naf}-literal of the form \text{naf} \( A \) in \( R \)’s body and \( A \in J \);
2. Delete every defeasible rule of the form
\[
(\text{\text{naf}} \ R \ L : \ \text{-Body}) \in Q
\]
such that \text{defeated}(r, L) \in J.
3. Remove all \text{naf}-literals from the remaining rules.
4. Remove all tags from the remaining rules.

Note that \( \frac{\overline{Q}}{\overline{J}} \) is a normal (non-defeasible) logic program without \text{naf}.

When dealing with stable models, it is often assumed that interpretations are consistent [19]. All the definitions and results in this section extend to this case straightforwardly.

**Definition 12 (Stable model)** A Herbrand interpretation \( M \) is a stable model of a non-disjunctive aspda \( \mathcal{P} \) with respect to the argumentation theory \( \mathcal{AT} \), if \( M \) is a minimal Herbrand model of \( \frac{\overline{\mathcal{P} \cup \mathcal{AT}}}{\overline{\mathcal{S}}} \).

Note that \( \frac{\overline{\mathcal{P} \cup \mathcal{AT}}}{\overline{\mathcal{S}}} \) is a Horn logic program, so here minimal models are meant in the sense of the regular Horn logic, not in the sense of Definition [17].

The next theorem shows that non-disjunctive aspdas can be implemented using ordinary logic programming systems that support the stable model semantics (e.g., DLV [30]).

**Theorem 1 (Reduction for stable model semantics)**
Let \( \mathcal{P} \) be a non-disjunctive aspda and \( \mathcal{AT} \) an argumentation ruleset. Then the following two sets coincide:

- The set of stable models of \( \mathcal{P} \) with respect to \( \mathcal{AT} \).
- The set of stable models of the ordinary logic program \( \mathcal{P}' \cup \mathcal{AT}' \), where \( \mathcal{P}' \) is obtained from \( \mathcal{P} \) by converting every defeasible rule
\[
(\text{\text{naf}} \ R \ L : \ \text{-Body}) \in \mathcal{P}
\]
into the plain rule of the form
\[
L : \ \text{-Body} \ \land \ \text{naf} \ \text{defeated}(r, L)
\]
and removing all the remaining tags; and \( \mathcal{AT}' \) is obtained from \( \mathcal{AT} \) by simply removing all the tags.

**Proof:** Let \( \mathcal{S} \) be a Herbrand interpretation for \( \mathcal{P} \cup \mathcal{AT} \). According to Definition [11], \( \frac{\overline{\mathcal{P} \cup \mathcal{AT}}}{\overline{\mathcal{S}}} \) is obtained through the following steps:

1. Delete every rule \( R \in \mathcal{P} \cup \mathcal{AT} \) such that there is a \text{naf}-literal of the form \text{naf} \( A \) in \( R \)'s body and \( A \in \mathcal{S} \);
2. Delete every defeasible rule of the form
\[
(\text{\text{naf}} \ R \ L : \ \text{-Body}) \in \mathcal{P}
\]
such that \text{defeated}(r, L) \in \mathcal{J}.
3. Remove all \text{naf}-literals from the remaining rules.
4. Remove all tags from the remaining tagged rules.

Note that this makes \( \frac{\overline{\mathcal{P} \cup \mathcal{AT}}}{\overline{\mathcal{S}}} \) into an ordinary logic program.

According to the definition of \text{Quotient} in the ordinary stable model semantics [20], the quotient of \( \mathcal{P}' \cup \mathcal{AT}' \) by \( \mathcal{S} \), is obtained through the following steps:

1. Delete every rule \( R \in \mathcal{P}' \cup \mathcal{AT}' \) such that there is a \text{naf}-literal of the form \text{naf} \( A \) in \( R \)'s body and \( A \in \mathcal{S} \);
2. Remove all \text{naf}-literals from the remaining rules in \( \mathcal{P}' \cup \mathcal{AT}' \).

From the above it can be safely inferred that the ASPDA quotient \( \frac{\overline{\mathcal{P} \cup \mathcal{AT}}}{\overline{\mathcal{S}}} \) is the same set of ordinary logic rules as the (ordinary) quotient of \( \mathcal{P}' \cup \mathcal{AT}' \) by \( \mathcal{S} \). For instance, consider a defeasible rule \( @r \ L : \ \text{-Body} \in \mathcal{P} \). If \text{defeated}(r, L) \in \mathcal{S} \), this rule will be deleted by the process of construction of \( \frac{\overline{\mathcal{P} \cup \mathcal{AT}}}{\overline{\mathcal{S}}} \). The corresponding rule
\[
L : \ \text{-Body} \ \land \ \text{naf} \ \text{defeated}(r, L)
\]
will be deleted by the construction of the (ordinary) quotient of \( \mathcal{P}' \cup \mathcal{AT}' \) by \( \mathcal{S} \).

The claim now follows from the above and the definitions of stable models in ASPDA and in the classical case (Definition [12] and the one in [20]).

For rules with disjunctions in the head, stable models are called answer sets and we will now generalize the above semantics to such rules. In generalizing aspdas to defeasible rules, the main difficulty is to find an analog of the reduction theorem (Theorem [1]).

**Example 1** Consider a disjunctive program that has the following defeasible rules:
\[
@r1 \ a \lor b \lor c.
@r2 \ d \lor e.
\]

The ordinary stable models of this program are \{a, d\}, \{a, e\}, \{b, d\}, \{b, e\}, \{c, d\}, and \{c, e\}. Suppose now that the proposition \( a \) cannot be true when either \( a \) or \( e \) holds, and that \( b \) is also incompatible with \( c \). These constraints are specified as the following facts:
Suppose, in addition, that rule \( r_1 \) has a higher priority than \( r_2 \), which we specify using the fact

\[
\text{\#overrides}(r_1, r_2).
\]

Intuitively, \{a, d\}, \{a, e\}, and \{b, e\} can no longer be models due to the incompatibility constraints, while the models \{b, d\}, \{c, d\}, and \{c, e\} are still possible. At the same time, one might feel that \{a\} is also a suitable model because \( r_1 \) overrides \( r_2 \), the proposition \( a \) makes \( r_1 \) true, and \( a \) is incompatible with both heads in rule \( r_2 \).

As it turns out, \{a\} may or may not be a defeasible stable model—it all depends on the associated argumentation rules. It would be a stable model of our aspda if the argumentation ruleset had the following rule instances:

\[
\text{\#overrides}(r_1, r_2) \land \text{\#opposes}(a, d) \land a.
\]

The following definitions generalize Definition 12 to disjunctive aspdas and make the intuition behind Example 1 precise.

Definition 13 (ASPDA quotient, disjunctive case) Let \( \mathcal{Q} \) be a disjunctive aspda, and let \( J \) be a Herbrand interpretation for \( \mathcal{Q} \). We define the ASPDA quotient of \( \mathcal{Q} \) by \( J \), written as \( \mathcal{Q}J \), by the following sequence of steps:

1. Delete every rule \( R \in \mathcal{Q} \) that has a literal of the form \( \text{naf} \ A \) in \( R \)'s body where \( A \in J \);
2. For every defeasible rule of the form \( \text{\#r} \ L_1 \lor \ldots \lor L_n : - \text{Body} \) in \( \mathcal{Q} \), delete every \( L_i \) such that \( \text{\#defeated}(r, L_i) \in J \). If all the \( L_i \)'s are deleted, delete the entire rule.
3. Remove all \( \text{naf} \)-literals from the remaining rules.
4. Remove all tags from the remaining tagged rules.

\[\square\]

The analog of Theorem 1 is as follows.

Theorem 2 (Reduction for the answer-set semantics) Let \( \mathcal{P} \) be a (disjunctive) aspda and \( AT \) an argumentation theory. Then the following two sets coincide:

- The set of answer sets for the aspda \( \mathcal{P} \) with respect to \( AT \).
- The set of answer sets for the ordinary logic program \( \mathcal{P}' \cup AT' \), where

  - \( \mathcal{P}' \) is obtained from \( \mathcal{P} \) by converting every defeasible rule
    \( \text{\#r} L_1 \lor \ldots \lor L_n : - \text{Body} \in \mathcal{P} \)

    into a collection of plain rules of the form

    \[
    \nu_{i \in K} L_i : - \text{Body} \land \bigwedge_{i \in K} \text{naf} \text{\#defeated}(r, L_i) \land \bigwedge_{j \in N \setminus K} \text{\#defeated}(r, L_j)
    \]

    for each non-empty subset \( K \subseteq N \), where \( N = \{1, \ldots, n\} \).

  - removing all the remaining tags.

- \( AT' \) is obtained from \( AT \) by simply removing all the tags.

\[\square\]

Proof: Let \( S \) be a Herbrand interpretation of \( \mathcal{P}' \cup AT \). By Definition 13 \( \mathcal{P}' = P \cup AT \) is constructed by the following steps:

1. Delete every rule \( R \in \mathcal{P}' \cup AT \) that has a literal of the form \( \text{naf} \ A \) in \( R \)'s body, where \( A \in S \);
2. For every defeasible rule of the form \( \text{\#r} L_1 \lor \ldots \lor L_n : - \text{Body} \in \mathcal{P} \cup AT \), delete every \( L_i \) such that \( \text{\#defeated}(r, L_i) \in S \). If all the \( L_i \)'s are deleted, delete the entire rule.
3. Remove all \( \text{naf} \)-literals from the remaining rules.
4. Remove all tags from the remaining tagged rules.

Note that \( \mathcal{P} = P \cup AT \) is an ordinary disjunctive logic program. For future reference, let us denote it \( Q_1 \).

By definition of the quotient in the ordinary answer set semantics [19], the quotient of \( \mathcal{P}' \cup AT' \) by \( S \), is obtained from \( \mathcal{P}' \cup AT' \) by these steps:

(i) Delete every rule \( R \in \mathcal{P}' \cup AT' \) that has a literal of the form \( \text{naf} \ A \) in \( R \)'s body, where \( A \in S \);
(ii) Remove all \( \text{naf} \)-literals from the remaining rules.

Let us denote the resulting logic program with \( Q_2 \). We will call the rules in \( Q_1 \) and \( Q_2 \) the reducts of the original rules in \( \mathcal{P} \cup AT \) and \( \mathcal{P}' \cup AT' \), respectively.

Now consider a rule \( R (\text{\#r} L_1 \lor \ldots \lor L_n : - \text{Body}) \in \mathcal{P} \).
If there is a literal of the form $\mathsf{naf} A$ in $R'$’s body and $A \in S$, $R$ would be deleted and its reduct will be neither in $Q_1$ nor $Q_2$.

If no such literal $\mathsf{naf} A$ exists in $R$ then $\text{Body}$ of the reduct of $R$ does not contain $\mathsf{naf}$-literals. Let $K_0 \subseteq \{1, \ldots, n\}$ be a subset such that $S \models \mathsf{naf} \mathsf{defeated}(r, L_i)$, for all $i \in K_0$, and $S \not\models \mathsf{defeated}(r, L_j)$, for all $j \not\in K_0$. Then, if $K_0 \neq \emptyset$,

- $Q_1$ would contain the rule $\lor_{K_0} L_i : \neg \text{Body} \lor \mathsf{naf} \mathsf{defeated}(r, L_i)$, for each non-empty subset $K \subseteq N = \{1, \ldots, n\}$.
- $P' \cup AT'$ would contain a set of rules of the form

$$\lor_{K} L_i : \neg \text{Body} \lor \mathsf{naf} \mathsf{defeated}(r, L_i) \lor \mathsf{defeated}(r, L_j) \lor \lor_{N - K} \mathsf{defeated}(r, L_j)$$

for each non-empty subset $K \subseteq N = \{1, \ldots, n\}$. During the construction of $Q_2$, after step (i) the only remaining rules will be of the form

$$\lor_{K} L_i : \neg \text{Body} \lor \mathsf{naf} \mathsf{defeated}(r, L_i) \lor \mathsf{defeated}(r, L_j) \lor \lor_{N - K} \mathsf{defeated}(r, L_j)$$

for each $K$ such that $K \subseteq K_0$. After step (ii), the reducts of $R$ that will remain in $Q_2$ would be:

$$\lor_{K} L_i : \neg \text{Body} \lor \mathsf{naf} \mathsf{defeated}(r, L_i) \lor \mathsf{defeated}(r, L_j) \lor \lor_{N - K} \mathsf{defeated}(r, L_j)$$

for each $K$ such that $K \subseteq K_0$. Among these rules, only one rule, $\lor_{K_0} L_i : \neg \text{Body} \lor \mathsf{naf} \mathsf{defeated}(r, L_i)$, can possibly have a body entailed by $S$. Furthermore, $S$ entails this rule if and only if $S$ entails $\lor_{K_0} L_i : \neg \text{Body}$, which is a reduct of $R$ in $Q_1$.

If $K_0 = \emptyset$ then, for $i=1, \ldots, n$,

- $Q_1$ has no reducts of $R$. So, the entire rule is deleted in step 2 (of ASPDA quotient).

- $P' \cup AT'$ must contain the rules of the form

$$\lor_{i \in K} L_i : \neg \text{Body} \lor \lor_{i \in K} \mathsf{naf} \mathsf{defeated}(r, L_i) \lor \lor_{j \in N - K} \mathsf{defeated}(r, L_j)$$

for each non-empty subset $K \subseteq N = \{1, \ldots, n\}$. Each such rule contains at least one literal $\mathsf{naf} \mathsf{defeated}(r, L_i)$ in the rule body. Since $K_0 = \emptyset$ implies that all such literals are false in $S$, step (i) in the construction of $Q_2$ eliminates all the above rules. So, neither $Q_1$ nor $Q_2$ will have any reducts of $R$.

It can now be seen that $S$ is a minimal Herbrand model of $Q_1$ if and only if $S$ is a minimal Herbrand model of $Q_2$. In other words, $S$ is an answer set for the $\mathsf{aspda}$ $P$ with respect to $AT$ if and only if $S$ is an answer set for the ordinary logic program $P' \cup AT'$. $\square$

With this theorem, it is now straightforward to verify that the answer sets for the $\mathsf{aspda}$ in Example 1 are precisely as described there.

Theorem 2 shows that a reduction exists from ASPDA to ASP, but that particular reduction is exponential in size with respect to the original program. With a little more care, a polynomial reduction can be constructed, as has been recently shown by Faber [16].

### 3.3. Reduction to the Non-Disjunctive Case

In ordinary answer-set programming, some disjunctive rules can be reduced to the non-disjunctive case via the so-called shifting transformation. This transformation would replace the rule $L_1 \lor \ldots \lor L_n : \neg \text{Body}$ with $n$ new rules

$$L_i : \neg \text{Body} \lor \lor_{1 \leq j \leq n, j \neq i} \mathsf{naf} L_j$$

(2)

where $1 \leq i \leq n$. We will use $\mathsf{shift}(P)$ to denote such transformation of a (non-defeasible) disjunctive logic program. For example, consider a program consisting of one rule $p \lor q \lor s : \neg \text{body}$, the shifting of the program is

$$p : \neg \text{body} \lor \mathsf{naf} q \lor \mathsf{naf} s.$$

$q : \neg \text{body} \lor \mathsf{naf} p \lor \mathsf{naf} s.$

$s : \neg \text{body} \lor \mathsf{naf} q \lor \mathsf{naf} p.$
Ben-Eliyahu and Dechter [4] have shown that the above shifting transformation is an equivalence transformation for so-called head-cycle free programs. We reproduce that definition below adjusting it for disjunctive aspdas.

**Definition 15** [4] The dependency graph $G_P$, of a ground aspda $P$, is a directed graph where nodes are ground literals. An edge going from literal $L$ to literal $L'$ exists if and only if there is a rule in which $L$ appears positively in the body and $L'$ is a head literal. An aspda is head-cycle free if and only if its dependency graph does not contain directed cycles that connect literals belonging to the head of the same rule. □

In the above example, if $p$, $q$, $s$ have only negative occurrences (or no occurrences at all) in body then the aspda consisting only of the rule

$$
\text{@r} p \lor q \lor s : \text{– body}
$$

is head-cycle free.

Under certain restrictions, the head-cycle free property for $P \cup AT$ can be reduced to head-cycle freedom for $P$. For example, if the literals that appear in rule heads in $AT$ do not appear in any rule body in $P$, and $AT$ is non-disjunctive, then $P \cup AT$ is head-cycle free if and only if $P$ is head-cycle free. This is satisfied in the argumentation ruleset $AT^{DL}$ in Section 4.2. It is also satisfied in the argumentation ruleset $AT^{AGCLP}$ in Section 4.3 if `overrides` and `opposes` do not appear in rule bodies in $P$ (which normally is the case).

An interesting question is whether a shifting transformation analogous to ordinary answer-set programming exists, and an equivalence result holds for disjunctive aspdas.

**Definition 16** Let $P$ be a disjunctive aspda. We define t-shifting of $P$, $t\_shift(P)$, as a non-disjunctive aspda obtained from $P$ by replacing each rule of the form $(\text{@r} L_1 \lor \ldots \lor L_n : \text{– Body}) \in P$ with $n$ new rules

$$
\text{@r} L_i : \text{– Body} \land \bigwedge_{1 \leq j \leq n, j \neq i} \text{naf} L_j
$$

where $1 \leq i \leq n$. □

Surprisingly, it turns out that $t\_shift(P)$ is not equivalent to $P$ even for head-cycle free aspdas. To see this, consider the following rule set, $P^{ex}$:

$$
\text{@r1} a \lor b \lor c.
\text{@r2} d.
\text{@r3} c.
$$

Suppose that the associated argumentation rules imply $\text{defeated}(r, c)$ and does not imply any other $\text{defeated}(\ldots)$ facts involving the above rules. Then $P^{ex}$ would have the following answer sets: {a, d, c} and {b, d, c}. In contrast, the above t-shifting transformation yields the following non-disjunctive aspda, $t\_shift(P^{ex})$:

$$
\text{@r1} a : \text{– naf} b \land \text{naf} c.
\text{@r2} b : \text{– naf} a \land \text{naf} c.
\text{@r1} c : \text{– naf} a \land \text{naf} b.
\text{@r2} d.
\text{@r3} c.
$$

which has only one answer set: {d, c} with respect to the argumentation ruleset. This shows that $t\_shift$ is not an equivalence transformation under ASPDA.

Fortunately, a result similar to Ben-Eliyahu and Dechter’s does hold for disjunctive aspdas, but for a slightly different shifting transformation.

**Definition 17** The ASPDA shifting of an aspda $P$, written as $aspda\_shift(P)$, is a non-disjunctive aspda obtained from $P$ by replacing each strict rule with its t-shifting and replacing each defeasible rule of the form $(\text{@r} L_1 \lor \ldots \lor L_n : \text{– Body}) \in P$ with $n$ new defeasible rules and $2n$ new strict rules as follows:

$$
\text{@r} L_1 : \text{– Body} \land \bigwedge_{1 \leq j \leq n, j \neq i} \text{lit}(r, L_j).
\text{lit}(r, L_i) : \text{– nafl}_i.
\text{lit}(r, L_i) : = \text{defeated}(r, L_i).
$$

where $1 \leq i \leq n$. Here $\text{lit}(r, L_i)$, $1 \leq i \leq n$, are literals of the form $\text{newsym}_i([\text{Vars}_i])$, where $\text{newsym}_i$ is a fresh predicate name that depends only on $r$ and $L_i$, while $\text{Vars}_i$, the argument vector of the literal, is a vector of variables that occur in $r$ and $L_i$. We omit the rule tags for strict rules here. □

**Theorem 3** Let $P$ be an aspda and let $AT$ be an argumentation ruleset such that $P \cup AT$ is head-cycle free. There is a one-to-one relationship between the answer sets of $P$ with respect to $AT$ and the answer sets of $aspda\_shift(P)$ with respect to $t\_shift(AT)$. Namely, a Herbrand interpretation $S$ is an answer set of $P$ with respect to $AT$ if and only if $f(S)$ is an answer set of $aspda\_shift(P)$ with respect to $t\_shift(AT)$, where $f(S) = S \cup \{\text{lit}(r, L) \mid P \text{ contains a rule with}$

---

3 The works [13,21] developed similar shifting techniques.
tag \( r \) and with \( L \) in its head (possibly as a disjunct), so that either \( \text{defeated}(r, L) \in S \) or \( L \notin S \).

**Proof:** The proof consists of establishing five equivalences, which we denote \( \Leftarrow_1, ..., \Leftarrow_5 \).

\( S \) is an answer set of \( \mathcal{P} \) with respect to \( \mathcal{AT} \)

\( \Leftarrow_1 \)

\( S \) is a minimal Herbrand model of \( \mathcal{Q}_1 = \frac{\mathcal{P} \cup \mathcal{AT}}{S} \)

\( \Leftarrow_2 \)

\( S \) is an answer set of

\( \mathcal{Q}_2 = \text{shift}(\mathcal{Q}_1) = \text{shift}(\frac{\mathcal{P} \cup \mathcal{AT}}{S}) \)

\( \Leftarrow_3 \)

\( \mathcal{Q}_3 = \frac{\mathcal{Q}_2}{S} = \text{shift}(\frac{\mathcal{P} \cup \mathcal{AT}}{S}) \)

\( \Leftarrow_4 \)

\( f(S) \) is a minimal Herbrand model of

\( \mathcal{Q}_4 = \frac{\text{aspda}_\text{shift}(\mathcal{P}) \cup t\_\text{shift}(\mathcal{AT})}{f(S)} \)

\( \Leftarrow_5 \)

\( f(S) \) is an answer set of \( \text{aspda}_\text{shift}(\mathcal{P}) \) with respect to \( t\_\text{shift}(\mathcal{AT}) \).

In proving each equivalence, we will choose an arbitrary defeasible rule \( \mathcal{R} \) of the form \( \forall r \quad L_1 \lor \ldots \lor L_n : \text{Body} \in \mathcal{P} \) and an arbitrary strict rule \( \mathcal{T} \in \mathcal{AT} \), and then look at what happens to these rules after applying the quotient and shifting transformations to them. As in the proof of Theorem 2, we can assume that the bodies of \( \mathcal{R} \) and \( \mathcal{T} \) do not contain \( \text{naf} \)-literals (they are evaluated away in the quotients on both sides).

Let \( \mathcal{K}_0 \) be \( \{ k \mid S \models \text{naf} \text{defeated}(r, L_k), 1 \leq k \leq n \} \) and let \( \mathcal{K}_1 \) be \( \{ k \mid L_k \in S, 1 \leq k \leq n \} \).

\((\Leftarrow_1)\): This follows from Definition 14

\((\Leftarrow_2)\): By definition, every defeasible rule \( \mathcal{R} \in \mathcal{P} \) gives rise to the following single rule \( \mathcal{R}_1 \) in the quotient \( \mathcal{Q}_1 \):

\[
\bigvee_{i \in \mathcal{K}_0} L_i : \text{Body}
\]

(4)

If \( |\mathcal{K}_0| = 0 \), \( \mathcal{R} \) gives rise to no rule.

The strict rules \( \mathcal{T} \in \mathcal{P} \cup \mathcal{AT} \) give rise to \( \mathcal{T}_1 \) in \( \mathcal{Q}_1 \) where \( \mathcal{T}_1 \) has the same head and body as \( \mathcal{T} \) but the tag is stripped off. By definition, all rules in \( \mathcal{Q}_1 \) are either \( \mathcal{R}_1 \)s or \( \mathcal{T}_1 \)s and are obtained in the above way. So, \( \mathcal{Q}_1 \) consists of the rules of the form (4) or of the strict rules from \( \mathcal{P} \cup \mathcal{AT} \) that lost their tag.

\( \mathcal{Q}_2 \) is constructed from \( \mathcal{Q}_1 \) via shifting of ordinary (non-defeasible) disjunctive rules. A rule \( \mathcal{R}_1 \) of the form (4) produces \( |\mathcal{K}_1| \) rules of the form

\[
L_i : \text{Body} \land \bigwedge_{j \in \mathcal{K}_0, j \neq i} \text{naf} L_j \quad (5)
\]

for \( i \in \mathcal{K}_0 \). The strict rule \( \mathcal{T}_1 \in \mathcal{Q}_1 \) gives rise to the rules \( \text{shift}(\mathcal{T}_1) \).

Since, by assumption, \( \mathcal{Q}_1 \) does not contain \( \text{naf} \)-literals, \( \mathcal{S} \) is a minimal Herbrand model of \( \mathcal{Q}_1 \) iff \( \mathcal{S} \) is an answer set of \( \mathcal{Q}_1 \). Observe that:

- \( \mathcal{Q}_3 \) is an ordinary (non-defeasible) disjunctive logic program,
- \( \mathcal{P} \cup \mathcal{AT} \) is head-cycle free, so \( \mathcal{Q}_1 = \frac{\mathcal{P} \cup \mathcal{AT}}{S} \) is head-cycle free,

Therefore, as shown in [4,13,21], \( \mathcal{S} \) is an answer set of \( \mathcal{Q}_1 \) iff \( \mathcal{S} \) is an answer set of \( \mathcal{Q}_2 = \text{shift}(\mathcal{Q}_1) \).

\( \mathcal{Q}_2 \) contains no rules other than those in (5) and \( \text{shift}(\mathcal{T}_1) \).

\((\Leftarrow_3)\): \( \mathcal{Q}_3 = \frac{\mathcal{Q}_2}{S} \) is constructed according to Definition 13. Strict rules in \( \mathcal{Q}_3 \) all have the form \( \text{shift}(\mathcal{T}_1) \) and defeasible rules are obtained as follows:

3-a. If \( |\mathcal{K}_1 \cap \mathcal{K}_0| \geq 2 \), the rules of the form (5) yield nothing in \( \mathcal{Q}_3 \). Indeed, for each rule in (5), there must exist at least one \( j \) satisfying \( j \in \mathcal{K}_0, j \neq i \), and \( L_j \in S \), so every such rule will be deleted after Step 1 in Definition 13.

3-b. If \( |\mathcal{K}_1 \cap \mathcal{K}_0| = 1 \), (5) yields \( \{ L_i : \text{Body} \mid i \in \mathcal{K}_1 \cap \mathcal{K}_0 \} \) in \( \mathcal{Q}_3 \). This is because every rule in (5) such that \( i \notin \mathcal{K}_1 \) is deleted in Step 1 in Definition 13 and the rules such that \( i \notin \mathcal{K}_0 \) are deleted in Step 2. The \( \text{naf} \)-literals in the remaining rule are deleted in Step 3.

3-c. If \( |\mathcal{K}_1 \cap \mathcal{K}_0| = 0 \), (5) yields the rules \( \{ L_i : \text{Body} \mid i \in \mathcal{K}_0 \} \) in \( \mathcal{Q}_3 \). This is because the rules in (5) such that \( i \notin \mathcal{K}_0 \) are deleted in Step 2 of Definition 13 while the \( \text{naf} \)-literals in the remaining rules are deleted in Step 3.

\((\Leftarrow_5)\): The fifth equivalence in the proof of the theorem is a direct consequence of Definition 14, so we dispense with this case before the fourth equivalence.
Each strict rule for each of the form (3) in $f_1$ of $Q_2$, and $Q_{4}^{4}$ which consists of rules obtained from the strict rules in $Q_2$.

$Q_4$ is divided into $Q_{4}^{4}$ and $Q_{4}^{4}$ the same way. $Q_4$ is obtained from $aspd_addshift(P) \cup t_shift(AT)$ by applying the steps in Definition $[11]$. Recall that each defeasible rule $R \in P$ gives rise to a collection of rules of the form $[3]$ in $aspd_addshift(P)$, which in $Q_4$ become

$$L_4 := \text{Body} \land \bigwedge_{1 \leq j \leq n, j \neq 1} \text{lit}(r, L_4).$$

(6)

for each $i \in K_0$;

$$\text{lit}(r, L_4) := \$\text{defeated}(r, L_4).$$

(7)

for each $1 \leq j \leq n$;

$$\text{lit}(r, L_4).$$

(8)

for each $j \notin K_1$. All these rules constitute $Q_{4}^{4}$. Each strict rule $T \in P \cup AT$ gives rise to the set of rules $t_{shift}(T)$ in $Q_4$. These rules constitute $Q_{4}^{4}$.

The difference between $t_{shift}(T)$ (used in $Q_4$) and $shift(T)$ used in $Q_3$ is that $T$ has a tag while $T_1$ does not. The quotient operation removes tags, so $t_{shift}(T) = shift(T)$ for some $T$ and every rule in $Q_{4}^{4}$ has the form $shift(T)$ for some $T_1$ (which is obtained from $T$ by tag removal), we have:

$$f(S)$$ is a Herbrand model of $Q_{4}^{4}$ iff $S$ is a Herbrand model of $Q_{4}^{4}$.

(9)

Now consider the defeasible rules in $Q_{4}^{4}$. Since $[7]$ and $[8]$ are the only rules that define $\text{lit}(r, L_4)$, it follows that $f(S) \not\models \text{lit}(r, L_4)$, $\forall j \in K_1 \cap K_0$, and $f(S) \models \text{lit}(r, L_4)$, $\forall j \notin K_1 \cap K_0$. Under $f(S)$,

4-a. If $|K_1 \cap K_0| \geq 2$, the rules in $[6]$ yield nothing in $Q_4$, since for each $i \in K_0$, every rule $[6]$ has some body literal $\text{lit}(r, L_4)$ such that $f(S) \not\models \text{lit}(r, L_4)$.

4-b. If $|K_1 \cap K_0| = 1$, $[6]$ gives rise to a single rule $L_4 := \text{Body}$, where $i \in K_1 \cap K_0$. Indeed, any rule in $[6]$ such that $i \notin K_1 \cap K_0$ has some body literal $\text{lit}(r, L_4)$ such that $f(S) \not\models \text{lit}(r, L_4)$.

4-c. If $|K_1 \cap K_0| = 0$, $[6]$ gives rise to the following rules $L_4 := \text{Body}$, which is obtained by the same argument as before.

Every rule in $Q_{4}^{4}$ comes from $[6]$ or $[7]$ or $[8]$ and all the rules of the form $[7]$ and $[8]$ are satisfied in $f(S)$. By comparing (3-a),(3-b),(3-c) with (4-a), (4-b), (4-c), it follows that

$$f(S)$$ is a Herbrand model of $Q_{4}^{4}$ iff

$$S$$ is a Herbrand model of $Q_{4}^{4}$.

(10)

From (10) and (9) we obtain

$$f(S)$$ is a Herbrand model of $Q_4$ iff

$S$ is a Herbrand model of $Q_3$.

(11)

To complete the proof for the equivalence $\Leftrightarrow 4$, it remains to show that $f(S)$ is a minimal model of $Q_4$ if and only if so is $S$ for $Q_3$.

Minimality of $f(S)$: if $S$ is a minimal Herbrand model of $Q_3$, then $\forall A \in f(S), f(S) \rightarrow \{A\}$ cannot be a Herbrand model of $Q_4$ because:

- if $A \in S$, the minimality of $S$ for $Q_3$ implies that there must be a rule $R_1$ of the form $[5]$ or $shift(T)$ such that $S \not\models \{A\}$, By the previously established correspondence between the rules in $Q_3$ and $Q_4$, there is a rule $R_3 \in Q_4$ of the form $[6]$ or $t_{shift}(T)$, which, by construction, must be such that $f(S) \rightarrow \{A\}$.

- if $A = \text{lit}(r, L)$ for some $r$ and $L$, there must be some rule $R$ of the form (4-b) or (4-c) such that $f(S) \rightarrow \{A\}$, so $f(S)$ also cannot be a model of $Q_4$ in this case.

Minimality of $S$: If $f(S)$ is a minimal Herbrand model of $Q_4$, then for any $A \in S, S \rightarrow \{A\}$ cannot be a Herbrand model of $Q_3$. If it were a model then, by (11), $f(S \rightarrow \{A\}) \cap f(S)$ must be a Herbrand model of $Q_4$, contrary to the assumption that $f(S)$ is a minimal Herbrand model of $Q_4$.

This concludes the proof of $\Leftrightarrow 4$ and of the theorem.

4. Examples of Argumentation Rules

We will now introduce two very different sets of argumentation rules and then discuss how the choose of an argumentation ruleset affects the semantics on a number of simple knowledge bases.
4.1. A-GCLP [24,40]

Our first example is an ASPDA counterpart for the argumentation theory proposed in [40], which captures generalized courteous logic programs [24] under the well-founded semantics [13]. We will call this theory A-GCLP and will denote it by $\text{AT}^\text{AGCLP}$. It is this argumentation rule-set that was tacitly assumed in all the earlier examples in this paper.

In $\text{AT}^\text{AGCLP}$, the predicate $\text{defeated}$, which plays a key role in the semantics of aspdas, is defined in terms of the predicates $\text{opposes}$ and $\text{overrides}$. These predicates are defined by the knowledge engineer within the knowledge base via background axioms. In our case, $\text{AT}^\text{AGCLP}$ supplies the following background axioms for $\text{opposes}$:

\[
\text{opposes}(L_1, L_2) :\neg \text{opposes}(L_2, L_1).
\]

The auxiliary predicate $\text{defeats}$ is defined as follows:

\[
\text{defeats}(T, L) :\neg \text{defeats}(T', L', T, L).
\]

The predicate $\text{strict}$ is used here to distinguish strict rules from the defeasible ones. The predicate $\text{refutes}$ indicates when one rule refutes another. Refutation of a rule means that a higher-priority rule implies a conclusion that is incompatible with the conclusion implied by the first rule. This is defined as follows:

\[
\text{refutes}(T_1, L_1, T_2, L_2) :\neg \text{refutes}(T_1, L_1, T_2, L_2) \land \neg \text{defeated}(T_1, L_1) \land \neg \text{strict}(T_2, L_2).
\]

The definition of a conflict between two rules, represented by the predicate $\text{conflict}$ above, relies in turn on the notion of a candidate. A candidate rule-instance is one whose body is true in the knowledge base:

\[
\text{candidate}(T, L) :\neg \text{body}(T, L, B) \land B.
\]

Here the meta-predicate $\text{body}$ binds $B$ to the body of a rule with the tag $T$ and head $L$.

Conflicting rules are now defined as follows: rules are in conflict if they are both candidates and the literals in them are incompatible:

\[
\text{conflict}(T_1, L_1, T_2, L_2) :\neg \text{candidate}(T_1, L_1) \land \text{candidate}(T_2, L_2) \land \text{opposes}(L_1, L_2).
\]

Recall that the $\text{opposes}$ information is supplied by the knowledge engineer. However, argumentation rules may include additional background axioms. In our case, $\text{AT}^\text{AGCLP}$ supplies the following background axioms for $\text{opposes}$:

\[
\text{opposes}(L_1, L_2) :\neg \text{opposes}(L_2, L_1).
\]

The first is a symmetry axiom that states that opposition is a reciprocal relation. The second axiom states that literals and their negations are in opposition to each other. The third axiom is a constraint that says that opposing literals cannot be both true in the same possible world.

The relation $\text{overrides}$ is also mostly defined by the knowledge engineer. However, $\text{AT}^\text{AGCLP}$ also supplies a background axiom that establishes preference for strict rules over defeasible ones:

\[
\text{overrides}(T_1, L_1, T_2, L_2) :\neg \text{strict}(T_1, L_1) \land \neg \text{strict}(T_2, L_2).
\]

Owering is often specified via tags instead of tag-head pairs, and this was the form of overriding that we mostly used in the examples. The relationship between overriding through tag-head pairs and overriding via tags is defined by the following rule:

\[
\text{overrides}(T_1, L_1, T_2, L_2) :\neg \text{overrides}(T_1, L_1, T_2, L_2) \land \text{head}(T_1, L_1) \land \text{head}(T_2, L_2).
\]

Here $\text{head}$ is a meta-predicate that relates tags to the heads of the rules labeled with those tags. The body-occurrence of $\text{overrides}$ is the overriding relation over tags and the head occurrence is the overriding relation over tag-head pairs.

Similarly, $\text{strict}$ is also often specified over tags and the following axiom relates that to strictness at the
level of tag-head pairs.

\[ \#\text{strict}(T, L) := \#\text{strict}(T) \land \text{head}(T, L). \]

Having defined this argumentation ruleset precisely, we can now come back to Example [1] and verify that the \texttt{asnda} there has four answer sets as claimed: \{a\}, \{b, d\}, \{c, d\}, \{c, e\}.

### 4.2. Defeasible Logic [1]

Our second argumentation ruleset is intended to closely approximate Defeasible Logic under the stable model semantics as defined in [1, 3].

Defeasible Logic partitions all rules into strict, defeasible, and defeaters. The defeater rules are used only to defeat other rules, but they themselves do not produce any inferences. In our terms, this means that defeater rules are defeated defeasible rules whose only purpose is to block inferences produced by other rules. Strict and defeater rules are specified via the predicates \#\text{strict} and \#\text{defeater}. Other important restrictions in [1] are that it does not support disjunctions in the rule heads; opposition among literals is limited to \( p \) and \( \neg p \), for each \( p \); it does not use default negation, so all literals are \texttt{naf}-free; and the rule tags are also rule identifiers, so no two rules have the same tag. This implies that rule tags uniquely determine rule’s head and body and lets us simplify the argumentation rules by considering tags only and ignoring rule heads in most cases.

We can now formulate the argumentation rules, which we denote as \( AT^{DL} \), for Defeasible Logic under the stable model semantics as defined in [1].

\[ \$\text{defeated}(T, L) := \$\text{conflict}(T, T') \land \text{head}(T', L') \land \$\text{definitely}(L'). \]
\[ \$\text{defeated}(T, L) := \#\text{defeater}(T). \]
\[ \$\text{defeated}(T, L) := \$\text{overruled}(T). \]

Here \texttt{head} is a meta-predicate that binds \( L \) to the head of a rule with \texttt{Id} \( S \).

The predicate \$\text{definitely} is defined as follows:

\[ \$\text{definitely}(L) := \]
\[ \#\text{strict}(T) \land \text{head}(T, L) \land \text{body}(T, B) \land \text{each_definite}(B). \]

As in A-GCLP, \texttt{body} is a meta-predicate that binds \( B \) to the body of a rule with tag \( T \); \texttt{each_definite} is a meta-predicate; it is true when \$\text{definitely}(B) is true or when \( \text{body}(T, B) \) holds and \texttt{each_definite}(B) is thus true. In this way, facts provide the base case for the recursive definition of \$\text{definitely}(L).

The predicate \$\text{candidate} is defined as before except that it now depends only on rule tags rather than tags and heads:

\[ \$\text{candidate}(T) := \text{body}(T, B) \land \text{?B}. \]

It remains to define \$\text{overruled}, which relies on the notion of candidacy and conflict, as in \( AT^{AGCLP} \).

\[ \$\text{overruled}(T) := \]
\[ \$\text{conflict}(T, T') \land \$\text{candidate}(T') \land \text{naf} \$\text{refuted}(T'). \]
\[ \$\text{refuted}(T') := \]
\[ \$\text{conflict}(T, T') \land \$\text{candidate}(T') \land \$\text{overrules}(T, T') \land \text{naf} \#\text{defeater}(T'). \]
\[ \$\text{overrules}(T, T') := \]
\[ \text{head}(T', L) \land \text{head}(T', \text{neg} L). \]

At this point it is instructive to retrospect on the differences between the two sets of argumentation rules presented here. First, there are differences in syntax and in how priorities over the rules are specified:

1. \( AT^{DL} \) does not support \texttt{naf} or disjunction in rule heads;
2. \( AT^{AGCLP} \) is more general in that tags are not required to be distinct and inclusion of variables in the tags provides one more level of differentiation among rule instances.

The other main difference is in the way \$\text{defeated} is defined. In \( AT^{AGCLP} \), a rule \( ?S \) is defeated if it is overridden by another rule \( ?R \) such that that \( ?R \) conflicts with \( ?S \). In contrast, in \( AT^{DL} \), a rule \( ?T \) is defeated if it conflicts with a rule that is not overridden.
This leads to significant differences in the behavior of the two argumentation rule sets for the examples discussed in Section 4.3.

4.3. Examples

We now discuss a number of examples to help better understand the ASPDA semantics and the differences between the argumentation rulesets presented earlier. In all the examples, rules that have explicit tags are assumed to be defeasible and the rules without the tags are assumed strict.

Example 2 Consider again the turkey-shoot example presented in Section 2.

Under the $AT^{AGCLP}$ argumentation rules, this example set has two answer sets. One is
$$\{ \text{neg loaded(g1,1)}, \text{loaded(g2,1)}, \text{alive(3)} \}$$
and the other is
$$\{ \text{loaded(g1,1)}, \text{neg loaded(g2,1)}, \text{alive(3)} \}.$$  
Thus, $AT^{AGCLP}$ yields the expected result.

As to the $AT^{DL}$ argumentation rules and the logic in [1], this theory does not support disjunctions in rule heads directly. However, we can work around this issue by applying the shifting transformation. Shifting is applicable here because head-disjunctions in the turkey-shoot example are head-cycle-free. Under this transform, $AT^{DL}$ yields the same result as $AT^{AGCLP}$.  

Example 3 Figure 2 describes a scenario where a toxic discharge into a river caused massive reduction in fish population.

Here both $AT^{AGCLP}$ and $AT^{DL}$ yield the same conclusion:
$$\{ \text{fishCount(s0+1,$\text{Squamish}$,trout,400)}, \text{fishCount(s0+2,$\text{Squamish}$,trout,0)} \}$$
This is the expected result, meaning that up to the moment of the toxic discharge, the Squamish river had 400 trout and then all of them died.

Interestingly, the same conclusion would be reached under LPDA [40] —a sibling of ASPDA developed for the well-founded semantics—if we use either the very same argumentation ruleset $AT^{DL}$, which we used here, or $AT^{GCLP}$, a ruleset analogous to $AT^{AGCLP}$ but designed for the well-founded semantics [40].

Thus, in this example, both $AT^{AGCLP}$ and $AT^{DL}$ yield the same result and this is also true under the well-founded semantics.

Example 4 [40] Figure 3 specifies part of a game where blocks are moved from square to square on a board.

The argumentation rules $AT^{AGCLP}$ under ASPDA and $AT^{GCLP}$ under LPDA both give the same expected result in this case:
$$\{ \text{loc}(0, \text{block4}, \text{square7}), \text{loc}(1, \text{block4}, \text{square7}), \text{loc}(2, \text{block4}, \text{square7}), \text{loc}(3, \text{block4}, \text{square3}) \}.$$  

Again, $AT^{DL}$ does not handle this example directly, since the syntax in [1] does not include $\text{naf}$. However, [3] shows that $\text{naf}$ can be simulated using a transform that relies on $\text{neg}$ only. Under this transform, $AT^{DL}$ yields the same result as the other two theories.

Example 5 Figure 4 shows a scenario where a cycle exists in the $\text{#overrides}$ relation between a pair of opponents.

Under ASPDA, $AT^{DL}$ yields an answer set in which both $a$ and $b$ are true. Indeed, one can verify that the following literals are true:
$$\text{refuted}(r1,a), \text{refuted}(r2,b), \text{naf}\text{soverruled}(r1,a), \text{naf}\text{soverruled}(r2,b), \text{naf}\text{defeated}(r1,a), \text{naf}\text{defeated}(r2,b).$$

Hence this program has only one answer set, in which both $a$ and $b$ are true.

The intuition is: Each of the two rules $r1$ and $r2$ has some rule that overrides it, so both of them are refuted; since refuted rules cannot be used to overrule any other rule, there is no rule overruled, and there is no rule defeated.

However, $AT^{AGCLP}$ does not produce this answer set. Indeed, consider an interpretation in which both $a$ and $b$ are true. We can infer that
$$\text{refutes}(r1,a,r2,b), \text{refutes}(r2,b,r1,a)$$
are true, but
$$\text{naf defeated}(r1,a), \text{naf defeated}(r2,b)$$
cannot both be true. This shows that $a$ and $b$ cannot both be true. So $\{a,b\}$ is not an answer set. Instead, there are two answer sets: in one $a$ is true and in the other $b$ is true.

The intuition is: $r1$ and $r2$ refute each other; it cannot be decided which defeats the other, in one possible world $r1$ defeats $r2$ and $a$ is true, and in the other $r2$ defeats $r1$ and $b$ is true.

For the reader who is familiar with LPDA, which is based on the well-founded semantics, we will go through the same example under the argumentation rules $AT^{GCLP}$, a sibling of $AT^{AGCLP}$ mentioned earlier. The rules comprising $AT^{GCLP}$ let us draw the...
/* Initial facts, and an “exclusion” constraint that fish count has a unique value */
occupies(trout,Squamish).
fishCount(s0,Squamish,trout,400).

/* Action/event description that specifies causal change, i.e., effect on next state */
@event fishCount(?s+1,?r,?f,0) :- occurs(?s,toxicDischarge,?r) ∧ occupies(?f,?r).

/* Persistence (“frame”) axiom */
@frame fishCount(?s+1,?r,?f,?C) :- fishCount(?s,?r,?f,?C).

/* Action axiom has higher priority than frame axiom */
overides(event,frame).

/* An action instance occurs */
occurs(s0+1,toxicDischarge,Squamish).

/* moving a block from ?from to ?to, if ?to is free; after the move, ?from becomes free */
move(?s,?blk,?from,?to) :-
    moves(?s,?blk,?from,?to) ∧ loc(?s,?blk,?from) ∧ naf loc(?s,?,?to).

move neg loc(?s+1,?blk,?from) :-
    moves(?s,?blk,?from,?to) ∧ loc(?s,?blk,?from) ∧ naf loc(?s,?,?to).

/* frame axioms: location of a block keeps the same */
frame loc(?s+1,?blk,?pos) :- loc(?s,?blk,?pos).
frame neg loc(?s+1,?blk,?pos) :- neg loc(?s,?blk,?pos).

/* each location is free, by default */
dloc neg loc(?s,?blk,?pos).

/* no block can be in two places at once */
overides(loc(?s,?blk,?y),loc(?s,?blk,?z)) :- posn(?y) ∧ posn(?z) ∧ ?y != ?z.

/* move-action beats frame axioms; move & initial state beats default location */
overides(move,frame).
overides(move,dloc).
overides(frame,dloc).

/* Facts: 16 squares. */
posn(square1). posn(square2). ... posn(square16).

/* initial state */
state loc(0,block4,square7).
overides(state,dloc).

/* State 2: block4 moves from square7 to square3 */
move(2,block4,square7,square3).

/* Fig. 2. Fish die-off example */

/* Fig. 3. Block moving example */

/* Fig. 4. Cycle of overides */
@ r1 a.
@ r2 b.
overides (a,b).
overides (r1,r2).
overides (r2,r1).
following conclusions: 
\$\text{refutes}(r_1,a,r_2,b),
\$\text{refutes}(r_2,b,r_1,a),
\$\text{refuted}(r_1,a), \text{ and }
\$\text{refuted}(r_2,b).

It now follows that:
\$\text{defeats}(r_1,a,r_2,b),
\$\text{defeats}(r_2,b,r_1,a)
are true. Consequently, the rules with tag-head pairs 
r_1,a \text{ and } r_2,b \text{ are both defeated, so both } a \text{ and } b
are false in LPDA. This is somewhat in line with the
behavior we saw from \$A^\text{AGCLP} under ASPDA, but
differences should have been expected, since there is
always a unique well-founded model under LPDA. \[] 

Our last example illustrates the semantics of 
AS- 
PDA on a number of simple “edge” cases, which are
unlikely to be found in practice. The example shows 
that our semantics is quite reasonable even for such
unusual aspdas.

**Example 6** Let an aspda consist of only one rule:
\$r \ a.
We will look at this aspda under different argumentation rulesets.
With respect to the argumentation ruleset
\$\text{defeated}(r,a), \text{our aspda has one answer set where } \$\text{defeated}(r,a)
is true and a false.

For the argumentation ruleset
\$\text{defeated}(r,a) :- a.
the above aspda has no answer sets.
Finally, if the argumentation ruleset is
\$\text{defeated}(r,a) :- \neg \text{af} \ a.
then there are two answer sets:
- \$\text{defeated}(r,a) \text{ is true and } a \text{ is false.}
- a \text{ is true and } \$\text{defeated}(r,a) \text{ is false.} \[]

5. Comparison with Other Work

Although a great deal of work has been devoted to
various theories of defeasible reasoning, only a few
considered disjunctive information or tried to unify the
different frameworks for such reasoning. The notable exceptions are the works \[22][11][56][9], which had
goals similar to ours. Due to the large volume of liter-
ature on defeasible reasoning, we will focus on the
above works, since they are related to our work most
closely. We refer the reader to a survey \[12\] for a dis-
cussion of the various individual theories of defeasibil-
ity.

Defeasible reasoning with disjunctive information in
the propositional case was studied in \[6\]. Bucca-
furri et al. \[9\] introduced a variant of disjunctive logic
programming with inheritance, called \$DLP^<\text{c}$. A key
feature in such inheritance systems is overriding of the
inherited information by more specific information,
which can be viewed as a specialized form of de-
feasible reasoning. Nonmonotonic inheritance can be
represented by means of argumentation rules, although
we have not studied the extent to which this is possible in \$DLP^<\text{c}\$.

The logic of prioritized defaults \[22\] also does not
use the notion of argumentation rules, but it allows
for multiple theories of defaults for different applica-
tion domains. This is analogous to allowing argument-
ation rulesets to vary. However, defaults are defined
via meta-theories and the semantics in \[22\] is given by
meta-interpretation. What we call an “argumentation
ruleset” is implicit in the meta-interpreters, and no in-
dependent model theory is given. In contrast, our
approach abstracts all the differences between the various
theories for defaults to the notion of an argumentation
ruleset with a simple interface to the user-provided do-
main description, the predicate \$\text{defeated}. Our approach is model-theoretic and it covers both the well-
founded semantics \[40\] and answer sets (the present paper). It unifies the theories of Courteous Logic Pro-
gramming, Defeasible Logic, Prioritized Defaults, and
more.

Delgrande et al. \[11\] propose a framework for or-
dered logic programming, which can use a variety of
preference handling strategies. For each strategy, this
approach devises a transformation from ordered logic
programs to ordinary logic programs. Each transfor-
mation is custom-made for the particular preference-
handling strategy, and the approach was illustrated by
showing transformations for several strategies, includ-
ing two described in earlier works \[42][15\].

Unlike ASPDA, the framework of Delgrande et al.
does not come with a unifying model-theoretic semantics.
Instead, the definition of preferred answer sets dif-
fers from one preference-handling strategy to another.
One of the more important conceptual differences be-
tween our work and \[11\] has to do with the nature of the
variable parts of the two approaches. In our case,
the variable part is the argumentation ruleset, which is
a set of definitions for concepts that a human reasoner
might use to argue why certain conclusions are to be
defeated. In case of \[11\], the variable part is the trans-
formation, which encodes a fairly low-level mecha-
nism: the order of rule applications required to gener-
ate the preferred answer set.\[19\] It is also important to note that each program transformation in [11] needs a compiler that contains hundreds of lines of Prolog code, while our approach requires no new software, and each argumentation ruleset typically contains 20-30 rules.

Eiter et. al. [15] set out to unify approaches to defeasible reasoning. Specifically, they present an adaptable meta-interpreter, which can be designed to simulate the approaches described in [8,42] among others. This framework is not as flexible as ASPDA and is fundamentally different from it: while ASPDA captures the essence of other approaches via argumentation rules, [15] captures these approaches in a less direct way, with the help of meta-interpretation.

The term “argumentation theory” was used to denote concepts that are related but significantly different from those studied in the present paper [7,17,34]. In these works, argumentation theories refer to proofs or sets of supporting premises rather than to rules that specify the notion of defeasibility. The focus of [7] is non-monotonic logic in general, while [17] is a procedural approach to defeasible reasoning. It is unclear whether these approaches can be captured as argumentation rules in our framework.

Argumentation theories were also used in a number of more closely related papers [35,36,27,14]. The focus of these works is development of the actual concepts that argumentation theories operate with. For instance, [35] uses Default Logic [39] to formalize the notions of defeat, defensible arguments, etc. Our work has a different focus in that we develop a general semantics for defeasible reasoning rather than dwelling on particular approaches to argumentation. The different argumentation rulesets (such as those in Section 4) are examples of the application of our general theory of defeasibility. These examples rely on some of the concepts that are analogous to those developed in [35,14].

For instance, the rulesets presented in Section 4 rely on the notion of defeated arguments, although those notions are not exactly the ones in [35,14].

Although defeasibility for disjunctive logic programs has been considered in restricted settings before [69], to the best of our knowledge, the present paper is the only work that studies the semantics of such logic programs in a general way. Defeasible disjunctive rules should not be confused with disjunctive logic programs under the answer-set semantics, as the latter does not explicitly represent defeasibility as a high-level concept but rather encodes it via default negation, not unlike the reduction described in Theorem 2.

6. Conclusions

This paper developed a novel theory of defeasible disjunctive logic programming under the answer-set semantics. It is a companion to our earlier work which developed a general theory of defaults and defeasibility through argumentation rules but was based on the well-founded semantics. Apart from the model theoretic semantics, and the reduction theorems, we have shown that head-cycle free disjunctive defeasible programs can be reduced to non-disjunctive ones, which mirrors an analogous result for non-defeasible disjunctive rules with default negation. To illustrate the power of the proposed framework, we gave two examples of argumentation rulesets. One is an adaptation for stable models of generalized courteous argumentation rules given in [40] for well-founded models. This theory was used in most of the examples in this paper. The second argumentation ruleset was intended to show how ASPDA simulates other approaches to defeasible reasoning; in this case the defeasible logic of [11]. We gave a detailed analysis of the behavior of the two argumentation rulesets on a number of interesting examples and compared the results with the behavior that would have resulted if we used defeasibility under the well-founded semantics of [40].

References


Note that argumentation rules can also encode rule application orderings.


