Taking the A-chain: Strict and Defeasible Implication in Argumentation Frameworks*

Adam Zachary Wyner and **Trevor Bench-Capon**
University of Liverpool
Department of Computer Science
Ashton Building
Liverpool, L69 3BX, United Kingdom

Abstract

Abstract Argumentation Frameworks (AFs) have proved a very fruitful basis for the exploration of the semantics of acceptability of sets of arguments as well as related complexity questions. However, applications of AFs are often held to require that arguments be given a more concrete instantiation. One popular method of instantiation is to postulate a knowledge base comprising facts and rules (typically both strict and defeasible in a Defeasible Logic (DL)). Arguments will then be demonstrations of claims from this knowledge base, which often require subarguments to demonstrate intermediate claims. Such a method is natural and readily applicable to logic programming, but it has been observed to exhibit a number of problems since allowing nodes in the AF this degree of structure introduces additional relations among arguments. In this paper we offer a principled view on instantiating arguments which retains the appeal of AFs, allows reasoning with respect to knowledge bases, yet avoids these problems. We restrict what can appear as nodes in the AF to simple assertions of literals or rule names. In addition, the attack relation between literals and rule names is restricted. We can thus retain the appropriate level of abstraction of the nodes of the AF as well as identify arguments which represent demonstrations from the knowledge base as structures in the AF. In this way, structured arguments can be identified in our framework as chains of these simpler arguments. The relations between these structured arguments which give rise to problems are now handled properly in the terms of which parts of the chain attack each other. Moreover, defeasible inference can be distinguished from strict inference by considering the attack relations within these chains. By restricting nodes of AFs in this way, the theoretical and computational benefits can be retained, while the desired structured arguments are captured by structures superimposed on the AF. In this way, our formalism provides a bridge between a deductive theory and AFs.

*During the writing of this paper, the first author was supported by the Estrella Project (The European project for Standardized Transparent Representation in order to Extend Legal Accessibility (ESTRELLA, IST-2004-027655)). We also thank three anonymous reviewers and Sanjay Modgil for comments. Comments welcome. Errors rest with the authors.

Introduction

Argumentation frameworks (AFs) ((Dung 1995), (Bondarenko *et al.* 1997) (Bench-Capon 2003), (Caminada & Amgoud 2007), among others) have been proposed to address a range of *non-monotonic* reasoning problems. In AFs, arguments are first class objects in an attack relation. The attack relation succeeds unless the attacking argument is itself defeated. AFs can be represented as directed graphs in which arguments are *nodes* and attacks are *arcs* between the nodes. AFs provide a very clean acceptability semantics (Dunne & Bench-Capon 2002), but require that the arguments have no *internal* structure and that the attack relation is the *only* relation between them.

However, AF approaches are not adequate to account for a range of uses of argumentation where arguments are related to a knowledge base and reasoning uses the *strict* (SI) and *defeasible* (DI) inference rules of Defeasible Logic (DL) (Pollock 1995). AFs with *abstract* and *atomic* arguments and a single attack relation cannot represent the *logical* or *philosophical* sense of argument ((Wyner, Bench-Capon, & Atkinson 2008) and (Hitchcock 2007)) as a *sequence* comprised of premisses, the illative relation *therefore*, and a conclusion. Nor can AFs represent *premise defeat*, *rebuttal*, *undercutting*, nor *argument schemes* and associated *critical questions* (Walton 1996). To relate AFs with these aspects of argumentation, *Rule-based Argumentation* (RBA) systems have been proposed ((Prakken & Sartor 1997), (García & Simari 2004), (Governatori *et al.* 2004), and (Amgoud *et al.* 2004)). However, according to (Caminada & Amgoud 2007), these systems fail to meet the properties of inference systems (closure and consistency) and produce counter-intuitive results. This is because the proposals introduce ambiguity and unclarity about what *constitutes* an argument in such frameworks and introduce additional *relations* among arguments in an AF, compromising the benefits of the clean AF semantics.

In this paper, we extend the *expressivity* of the language of AFs without increasing its *complexity*, clarifying the role of arguments in AFs and supporting rule-based reasoning directly in the AF. Only *literals* and *rule names* are nodes of the AF. They stand in sorted *attack* relations. Inference rules are defined as *structures* of such nodes and relations in the AF along with conditions relative to Dungian extensions. Arguments in the sense found in logic and philosophy are *superimposed* on such an AF rather than being included as nodes so that the nodes remain abstract and atomic. With this, we gain the capacities of RBA without compromising the AF semantics or introducing new semantics. The language we provide is novel, *unifies* different approaches to RBA, maintains the semantics of AFs, whilst making them useful for instantiated reasoning directly in AFs.

The structure of the paper is as follows. We review the basics of AFs (Dung 1995). We then introduce our AF language in which *literals* and *rule names* are *nodes* in attack relations. Strict and Defeasible Implication rules (SI and DI) are presented as *structures* within the AF. An *argument* in the philosophical sense ((Walton 1996) and (Hitchcock 2007)) is a structure comprised of SI or DI rules, assertions, and a conclusion derived from the rules and assertions. Arguments of this sort can be *chained* together. We then provide examples along with our solutions to *benchmark problems*. Finally, we discuss how the analysis can be used to express related Non-Monotonic reasoning techniques such as *abduction*.

Review of AF

An *Argumentation Framework* (AF) is a language comprised of objects, relations, and definitions of auxiliary concepts. For our purposes, we take (Dung 1995) as the most abstract system. Since we want to clarify the notion of *argument* itself, we refer to the basic objects as *ANodes* and their relations as *AArcs*; indeed, we do not want to introduce presumptions about the properties of the objects.

Definition – AF is a tuple <ANode, AArc>:

- ANode is a set of *objects*, n_1,...,n_n.
- AArc is an *attack* relation between objects. AArc(n_1, n_2) reads as object n_1 attacks object n_2. We assume that no object attacks itself.

The relevant auxiliary definitions are as follows, where R, S are subsets of Node:

Definition – Acceptable, Admissible, and Extensions

- x $\in$ ANode is *acceptable with respect to* S if for $\forall$y $\in$ ANode where AArc(y, x), $\exists$z $\in$ S where AArc(z, y).
- S is *conflict-free* if $\neg\exists$y $\exists$x $\in$ S, x $\neq$ y and AArc(x,y).
- A conflict-free set S is *admissible* if $\forall$x $\in$ S, x is acceptable with respect to S.
- S is a *preferred extension* iff it is a *maximal* (w.r.t. set-inclusion) admissible set.
- A node x is *credulously accepted* if x is a member of some preferred extension.
- A node is *skeptically accepted* if x is a member of every preferred extension.

Relationship of AF to Defeasible Logic

While it is usually claimed that AFs represent defeasible logic (Bondarenko *et al.* 1997), RBAs in practice use two distinct levels of representation (Caminada & Amgoud 2007). At the logic level, literals (positive and negative atomic propositions) are represented along with strict and defeasible inference rules; this is essentially the *knowledge base*. The AF level takes logical expressions as input, "packages" the logical expressions as arguments, specifies their attack relations, then produces Dungian extensions as output. To determine what "follows" from the logic given the AF extensions, one must, in effect, "unpackage" the arguments. We provide examples later. Not only is the relation between the logic and the AF complex, but the meaning of "argument", "argument attack", "premise", and "claim" are obscured. In particular, new relationships among arguments are introduced: an argument can appear as the premise of a more complex argument; two arguments can be chained together, where the conclusion of one is the premise of another; a complex argument can use both strict and defeasible implication. In addition, where an argument with subarguments is attacked, it is not clear what is being attacked – the subargment, the premise, the implicational rule, or the claim. (Caminada & Amgoud 2007) claim that sets of arguments in an AF can be used to infer conclusions. But it is not necessarily so that the conclusion of every acceptable "argument" is itself acceptable, for instance, in VAFs (Bench-Capon 2003); the definitions of an AF with DI would have to be extended to yield just the right set of nodes. Yet, in an AF, sets of "arguments" are just that, and it is unclear how to define the appropriate notion of "claim" since AFs do not have the requisite structure.

The chief aim and main novelty of this paper, then, is to formally clarify the relationship of AFs to Defeasible Logic. To do so, we *directly* incorporate the logic level into the AF level by introducing logical arguments as *superimposed structures* on a standard AF. Our key proposals:

- AF nodes represent literals and rule labels. AF arcs represent attacks between literals or between literals and rules.
- SI and DI rules are *structures* in an AF.
- Specify premises, rules, and claims in terms of SI and DI.
- Represent *assertions* directly in the AF.
- Define *arguments* in terms of rules, assertions, and conclusions relative to an AF.

We use our analysis to:

- Eliminate the ontological obscurity of arguments.
- Account for two example problems in previous theories.
- Account for *consistency* and *closure*.

The Properties of Strict and Defeasible Implication

As we are representing strict and defeasible implication directly in terms of an AF, SI in an AF must satisfy the properties of *Contraposition*, *Strengthening of the Antecedent*, and *Transitivity*; DI should not satisfy any of these properties. ϕ and ψ are meta-variables over literals.

Properties of Strict Implication

a. $[[\phi \rightarrow \psi] \equiv [\neg\psi \rightarrow \neg\phi]$
b. $[\phi \rightarrow \psi]$ implies $[[\phi \wedge \rho] \rightarrow \psi]$
c. $[[\phi \rightarrow \psi] \wedge [\psi \rightarrow \rho]]$, then $[\phi \rightarrow \rho]$

Once we present the basic elements of our analysis, we consider each of these properties and how they can be satisfied in an AF.

Strict and Defeasible Rules in AFs

In a DL, we represent *strict* implication with $\rightarrow$ and *defeasible* implication with $\rightsquigarrow$. With either sort of rule, we reason with literals. Thus, we sort our nodes into literals and rules. The role of the rules in logic is primarily to *facilitate reasoning about the literals*; that is, we reason *about* the literals *using* the rules. In this light, we regard the rule nodes N_r as *auxiliary elements* in an AF along the lines as discussed in (Wyner & Bench-Capon 2008) and further below. In addition, rules are relevant for our discussion of non-monotonic logic and circumscription.

With these sorts, we have three sorts of attack relations: literals attacking literals, rules attacking literals, and literals attacking rules. In our structured representation of inference in an AF, we use these nodes and attack relations.

Definition – AF_{DL} is a tuple $<$ANode, AArc$>$:

- ANode is $N_l \cup N_r$, where we sort the nodes into those which are literals N_l and rules N_r where:
  N_l is the set of literals, positive and negative atomic propositions, $\{\phi_1,...,\phi_n\} \cup \{\neg\phi_1,...,\neg\phi_n\}$;
  N_r is the set of rule names $\{\rho 1,...,\rho n\}$.
- AArc is the *attack* relation between objects of sort ANode, AArc $\subseteq$ ($AArc_{ll} \cup AArc_{lr} \cup AArc_{rl}$) where:
  $AArc_{ll} \subseteq (N_l \times N_l)$; $AArc_{lr} \subseteq (N_l \times N_r)$; $AArc_{rl} \subseteq (N_r \times N_l)$.

For brevity, we have no attacks of rules on rules, which expresses *undercut*. We assume double-negation elimination ($\neg\phi \equiv \neg\neg\phi$).

We assume the set of nodes of an AF_{DL} contains positive and negative literals, ϕ and $\neg\phi$, which we refer to as *contraries*. These attack one another, but not other literals, so they never both hold within a conflict-free set of nodes of any AF_{DL}.

Definition – Attacks between Contraries Suppose AF_{DL} ζ with literals ζ_{N_l} and attack relation ζ_{AArc}.

a. $\forall\, \phi \in \zeta_{N_l}$, $\neg\phi \in \zeta_{N_l}$.
b. $\forall\, \phi, \neg\phi \in \zeta_{N_l}$ $<\phi, \neg\phi>$ and $<\neg\phi, \phi> \in \zeta_{AArc}$.
c. $\forall\, \phi, \psi \in \zeta_{N_l}$, $<\phi, \psi>$ or $<\psi, \phi> \in \zeta_{AArc}$ if and only if $\psi \equiv \neg\phi$.

For example, suppose AF_{DL} ζ^1 with $\zeta^1_{N_l}$ = $\{\phi_1, \phi_2, \phi_3, \neg\phi_1, \neg\phi_2\}$, then ζ^1_{AArc} must have attacks between contraries $\{<\phi_1, \neg\phi_1>, <\neg\phi_1, \phi_1>, <\phi_2, \neg\phi_2>, <\neg\phi_2, \phi_2>\}$. If we filter out all rule nodes from AF_{DL} and considering *only* literals and their attacks, *preferred extensions* are *maximal sets* of *consistent literals*. In this respect, preferred extensions can be understood as *possible worlds*; we could define related notions of *information growth* from subsets to supersets of consistent literals.

For brevity, we consider only rules along the lines of *logic programming* (i.e. Horn-clauses) where there are conjunctive antecedents and a claim is a literal. We define functions to translate from the logical representation into an AF.

In correlating a logical representation with a set-theoretic representation that is suitable for AFs, we have several *utility functions*. There are two sets of these functions. First, we have functions with respect to *conjunctive formulae*. Second, we have utility functions to *translate* from the logical form of implications into a structure in an AF. The utility functions for conjunctions are used in the utility functions for implication. We give the definitions, then discuss them with examples.

Definition – Utility Functions for Conjuncts

- ConjSet is a function from a conjunctive proposition of literals to the set of literals: ConjSet($[\psi_1 \wedge ,..., \wedge\, \psi_n]$) $=_{def}$ $\{ \phi \mid \phi = \psi_1 \vee ,..., \vee\, \phi = \psi_n\}$
- RecAtt is a function from a conjunctive proposition of literals to a mutual attack relation among them: RecAtt($[\psi_1 \wedge ,..., \wedge\, \psi_n]$) $=_{def}$ $\{<\phi, \neg\phi>, <\neg\phi, \phi> \mid \phi \in$ ConjSet($[\psi_1 \wedge ,..., \wedge\, \psi_n]$)$\}$
- RuleAtt is a function from a conjunctive proposition of literals and a rule to the negations of the literals which attack a rule: Suppose $r_x \in N_r$. RuleAtt($[\psi_1 \wedge ,..., \wedge\, \psi_n]$, r_x) $=_{def}$ $\{<\neg\phi, r_x> \mid \phi \in$ ConjSet($[\psi_1 \wedge ,..., \wedge\, \psi_n]$)$\}$

Suppose a logical formula $[[\phi_1 \wedge \phi_2] \rightarrow \psi]$. To represent the conjoined antecedent in an AF_{DL}, we have a function *ConjSet* the correlated set of literals $\{\phi_1, \phi_2\}$. By assumption, in an AF_{DL}, each of these literals attack its negation, which is given by *RecAtt*. In addition, as given below, the logical form $[[\phi_1 \wedge \phi_2] \rightarrow \psi]$ is *reified* as a string "$[[\phi_1 \wedge \phi_2] \rightarrow \psi]$", which is a node of the rule name sort in the AF. The function *RuleAtt* introduces attacks on the rule node by negations of literals, as also show below.

We have several utility functions which are used for both strict and defeasible implication.

Definition – Utility Functions for Implications Suppose AF_{DL} ζ with objects ζ_{ANode} = ($N_r \cup N_l$) and relation ζ_{AArc}.

- Suppose γ is *either* $[\psi_1 \wedge ,..., \wedge\, \psi_n] \rightarrow \phi$ *or* $[\psi_1 \wedge ,..., \wedge\, \psi_n] \rightsquigarrow \phi$.
- (PremissNodes(γ) $\cup$ ClaimNodes(γ)) $\subseteq \zeta_{ANode}$.
- AArcs(γ) $\subseteq \zeta_{AArc}$.
- RuleName(γ) $=_{def}$ μ, where μ of sort N_r and μ is:
  - "Rule: $[\psi_1 \wedge ,..., \wedge\, \psi_n]$ strictly implies ϕ" if γ is $[\psi_1 \wedge ,..., \wedge\, \psi_n] \rightarrow \phi$, or;
  - "Rule: $[\psi_1 \wedge ,..., \wedge\, \psi_n]$ defeasibly implies ϕ" if γ is $[\psi_1 \wedge ,..., \wedge\, \psi_n] \rightsquigarrow \phi$
- PremissNodes(γ) $=_{def}$ $\{\rho, \neg\rho \mid \rho \in$ ConjSet($[\psi_1 \wedge ,..., \wedge\, \psi_n]$)$\}$.
- ClaimNodes(γ) = $_{def}$ $\{\sigma, \neg\sigma \mid \sigma \in$ ConjSet(ϕ)$\}$.
- AntecedentNodes(γ)$=_{def}$ ConjSet($[\psi_1 \wedge ,..., \wedge\, \psi_n]$).
- ConclusionNode(γ)$=_{def}$ $\{\phi\}$.

Implications are *structured* within an AF_{DL}: we associate them with a rule name; and they are associated with sets of nodes of *premisses*, *claims*, *antecedents*, *conclusions*. Note that the terminology of *premisses* and so on is related to but distinct from the familiar uses in logic and argumentation.

We have a seeming redundancy in *premisses-antecedents* and *claims-conclusions*; however, the sets are distinct, for the sets of *premisses* and *claims* contain positive and correlated negative literals, while the sets of *antecedents* and *conclusions* contain only the literals from the logical form (i.e. the set of *antecedents* is a proper subset of the set of *premisses*, and the set of *conclusions* is a proper subset of the set of *claims*). The distinctions help in discussing the representations of implications in AFs and are crucial in defining chains of implications. We speak of the set of *conclusions* for a consistent manner of expression and also to allow for extensions in the future where the conclusion is not just a literal.

For example, suppose γ is the logical form for strict implication $[[\phi_1 \wedge \phi_2] \rightarrow \psi]$: the rule name is a string "Rule: $[[\phi_1 \wedge \phi_2] \rightarrow \psi]$"; the premiss set is $\{\phi_1, \phi_2, \neg\phi_1, \neg\phi_2\}$; the antecedent set is $\{\phi_1, \phi_2\}$; the claims set is $\{\psi, \neg\psi\}$; and the conclusion set is $\{\psi\}$. An example with defeasible implication is similar.

Using these utility functions, we can define models in an AF_{DL} in which hold strict and defeasible implication as well as strict and defeasible assertions. We first consider strict implication, providing the definitions and some discussion along with an example.

Definition – Strict Implication Rule (SI) Suppose AF_{DL} ζ with objects ζ_{ANode} = ($\text{N}_r \cup \text{N}_l$) and relation ζ_{AArc}.

- $\gamma = [\psi_1 \wedge ,..., \wedge\, \psi_n] \rightarrow \phi$, with rule name μ.
- AArcs(γ) $=_{def}$ RecAtt($[\psi_1 \wedge ,..., \wedge\, \psi_n]$) $\cup$ RuleAtt($[\psi_1 \wedge ,..., \wedge\, \psi_n], \mu$) $\cup$ $\{<\mu, \neg\phi>\}$ $\cup$ RecAtt(ϕ).
- Condition: $\forall\, \chi \in \wp(\zeta_{ANode})$, if AntecedentNodes($\gamma$) $\subseteq \chi$, (ConclusionNode(γ) $\cup\, \chi$) is *acceptable* with respect to ζ.

The logical form of a strict implication (SI) is expressed as a *structure* within an AF_{DL}. The utility functions associate it with a rule name, and set of nodes for *premisses*, *claims*, *antecedents*, *conclusions*. In the definition for SI, we have defined an attack relation along with a condition on the sets of nodes in the AF_{DL}. Note that in the attack relation for SI, we assume that negations of antecedent literals attack the rule node and that the rule node attacks the negation of the conclusion. The condition stipulates that in *every* conflict-free set of ζ_{ANode} of an AF_{DL} where the antecedents hold, the conclusion holds as well.

For example, suppose the logical form is $[[\phi_1 \wedge \phi_2] \rightarrow \psi]$: the premiss set is $\{\phi_1, \phi_2, \neg\phi_1, \neg\phi_2\}$; the antecedent set is $\{\phi_1, \phi_2\}$; the claims set is $\{\psi, \neg\psi\}$; the conclusion set is $\{\psi\}$; and the rule name is a string "Rule: $[[\phi_1 \wedge \phi_2] \rightarrow \psi]$". $\neg\phi_1$ and $\neg\phi_2$ both attack the rule name; the rule name attacks $\neg\psi$.

The definition for DI is:

Definition – Defeasible Implication Rule (DI) Suppose AF_{DL} ζ with objects ζ_{ANode} = ($\text{N}_r \cup \text{N}_l$) and relation ζ_{AArc}.

- $\gamma = [\psi_1 \wedge ,..., \wedge\, \psi_n] \rightsquigarrow \phi$, with rule name ν.
- AArcs(γ) $=_{def}$ RecAtt($[\psi_1 \wedge ,..., \wedge\, \psi_n]$) $\cup$ RuleAtt($[\psi_1 \wedge ,..., \wedge\, \psi_n], \nu$) $\cup$ $\{<\nu, \neg\phi>, <\neg\phi, \nu>\}$ $\cup$ RecAtt(ϕ).
- Condition: $\exists\, \chi \in \wp(\zeta_{ANode})$ where AntecedentNodes(γ) $\subseteq \chi$, and (ConclusionNode(γ) $\cup\, \chi$) is *acceptable* with respect to ζ.

The differences between SI and DI are:

- The rule name of an SI and a DI are different
- the rule node of SI *attacks, but is not attacked by* the negation of the consequent, while in DI, the rule node *attacks and is attacked by* the negation of the consequent
- The conditions are different:
  - For SI, the condition requires that conclusion node is *sceptically* accepted in every conflict-free extension in which antecedent nodes hold; in such a case, every attacker on the conclusion node must be defeated.
  - For DI, it requires that the conclusion node is *credulously* accepted in some conflict-free extension in which antecedent nodes hold; it allows that the negation of the conclusion node is *credulously* accepted in some *other* extension where the antecedent nodes hold.

The rules both guarantee consistency since no literal and its negation can form part of a conflict free set. The definition of SI (but not DI) implies *closure*: for every preferred extension in which the antecedent nodes hold, the conclusion node holds as well. The rules cannot introduce conflict since the conclusion must be acceptable to the set which contains the antecedents. Furthermore, we show below the transitivity of SI but not of DI.

We have an incompatibility constraint:

Definition – SI and DI Incompatibility

- For ψ_1 ,..., ψ_n and ϕ, there is no AF_{DL} where $[\psi_1 \wedge ,..., \wedge\, \psi_n] \rightarrow \phi$ and $[\psi_1 \wedge ,..., \wedge\, \psi_n] \rightsquigarrow \phi$ are both defined.

We have definitions for assertions:

Definition – Strict and Defeasible Assertion Suppose AF_{DL} ζ with objects ζ_{ANode} and $\psi \in \zeta_{ANode}$.

- Strict Assertion of ψ in ζ: SA(ψ, ζ) $=_{def}$ ψ is skeptically accepted in ζ.
- Defeasible Assertion of ψ in ζ: DA(ψ, ζ) $=_{def}$ ψ is credulously accepted in ζ.

Strict Assertion requires that if the node ψ is attacked by a node ϕ, then ϕ itself must be defeated; Defeasible Assertion has no such requirement.

Examples and Cases

We illustrate the implication structures and give cases relative to assertions. We show how our analysis provides a semantics to support the properties of contraposition and strengthening of the antecedent, which are required for SI, but not for DI. For the purposes of illustration, we assume AF_{DL} only has the nodes and arcs (indicated as *arrows*) as illustrated, so only what appears can be asserted. Note that "not" is the graphic representation of $\neg$, *P* and *Q* represent literals, and the rule name can be inferred from the structure, so we abbreviated them as "Rule 1", "Rule 2", and so on.

Strict Implication

Our first example is an SI $P \rightarrow Q$, which is represented as follows in an AF_{DL} as in Figure 1. "Rule1" is the string "P strictly implies Q".

$$\text{P} \longleftrightarrow \text{not P} \longrightarrow \text{Rule1} \longrightarrow \text{not Q} \longleftrightarrow \text{Q}$$

Figure 1: Strict Implication

With respect to Figure 1, we have four cases where we read the following as "an assertion (e.g. as in (a.) *P*) relative to SI implies a preferred extension (e.g. as in (a.) {P, Rule1, Q})".

a. P: {P, Rule1, Q}

b. ¬P: {¬P, Q}, {¬P, ¬Q}

c. Q: {P, Rule1, Q}, {¬P, Q}

d. ¬Q: {¬P, ¬Q}

The extensions correlate to those cases of material implication of Propositional Logic where the implication is true. Note that if $\neg Q$ is asserted as in (d.), then whatever attacks it, here "Rule1", must be attacked; this implies that $\neg P$ must hold. Consequently, there is *no* conflict-free extension in which $\neg Q$ and *P both* hold since in every extension where $\neg Q$ holds $\neg P$ must hold as well. In this interpretation, strict implication is *false* where there is conflict-free where the antecedent *P* and the negation of the consequent $\neg Q$ both hold. As stated earlier, we have not considered attacks on the rule *per se*. We have included the rule name in the extensions, however, we are primarily interested in the literals and could filter out the rule names as they are auxiliary nodes that are used in calculating the extensions.

To assert $\neg P$ constitutes *Premise Defeat*, while to assert $\neg Q$ constitutes *Rebuttal*. Thus, our analysis incorporates these Argumentation-theoretic notions directly.

Defeasible Implication

Consider a DI $P \rightsquigarrow Q$ and the same cases. "Rule2" is the string "P defeasibly implies Q".

$$\text{P} \longleftrightarrow \text{not P} \longrightarrow \text{Rule2} \longleftrightarrow \text{not Q} \longleftrightarrow \text{Q}$$

Figure 2: Defeasible Implication

With respect to Figure 2, we have the following four cases:

a. P: {P, Rule2, Q}, {P, ¬Q}

b. ¬P: {¬P, Q}, {¬P, ¬Q}

c. Q: {P, Rule2, Q}, {¬P, Q}

d. ¬Q: {¬P, ¬Q}, {P, ¬Q}

Here, the assertion of the antecedent *P* does not guarantee the consequent *Q* in every preferred extension. By the same token, assertion of the negation of the consequent $\neg Q$ does not guarantee the antecedent *P* in every preferred extension.

The differences in attacks and conditions between SI and DI account for the different extensions.

Properties

Let us consider the properties of *contraposition* and *strengthening of the antecedent* with respect to these structures.

Contraposition *Contraposition* must hold for SI but does not hold for DI. In SI, the assertion of $\neg Q$ gives rise to an extension where $\neg P$ holds, accounting for the semantic contraposition property of SI; in SI, the only way for *Rule1* to be defeated is where $\neg P$ holds, which in this AF can only occur where $\neg Q$ is strictly asserted. Moreover, in brief, the contraposed formula $\neg Q \rightarrow \neg P$ has the same preferred extensions in the same cases with respect to the literals. Thus, the analysis provides a *semantics* to support contraposition.

However, contraposition does not hold for DI. This follows since in DI (and not in SI) $\neg Q$ itself attacks *Rule2*. Therefore, we have a choice between $\neg P$ attacking the rule or *P* attacking $\neg P$. Unlike SI, where we had no such choice, we cannot justify eliminating *P*. Thus, in DI and given the assertion $\neg Q$, $\neg P$ is *credulously accepted*.

Strengthening the Antecedent We can strengthen the antecedent of SI. In addition to being a property of SI, such cases play a role in the discussion of the problems. Suppose the logical form is *[*[P ∧ Q] → R] and Rule6 is "P and Q strictly imply R".

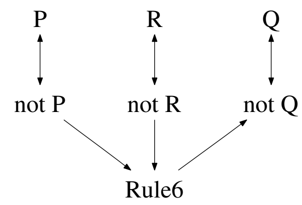

Figure 3: Strengthening the Antecedent

With respect to Figure 3, we have the following cases.

a. P, ¬R: {P, ¬R, ¬Q}, {P, ¬R, Q}

b. P, R: {P, R, Rule6, Q}

c. ¬Q: {P, ¬R, ¬Q}, {¬P, R, ¬Q}, {¬P, ¬R, ¬Q}

Examples (a.) and (b.) show the way conjunction works in SI. Example (c.) shows that we only *credulously accept* the antecedents where we have the contrapositive in an SI. This is important in the consideration of a problem below.

R-Chains and A-Chains

So far we have dealt with strict and defeasible implications in which there is no shared object. However, as implications can be *chained* together, we want to define their relationship. This is required for the analysis of *transitivity* of SI. We first provide the definitions, then examples and cases.

Following (Hitchcock 2007), we recursively define *arguments* as understood in logic and philosophy in terms of rules and assertions, which is in contrast to (Caminada & Amgoud 2007). We first recursively define *R(ule)-Chains* in terms of SI and DI rules. *A(rgument)-Chains*, which express the *philosophical sense of arguments*, are defined in terms of *R-Chains*.

In the base case, an R-Chain is either an SI or a DI. In the recursive case, the antecedents of the complex R-Chain are the union of the antecedents of the component R-Chains, the conclusion of the complex R-chain is the conclusion of one of the component R-Chains, the attack relation of the complex R-chain is the union of the attack relation of the component R-Chains, and the two component R-Chains must have overlapping claims and premisses.

Definition: R-Chains

- Base Case: R-Chain(γ,ζ), γ *is an R-Chain in* ζ, where ζ is an AF$_{DL}$ and γ is either an SI or a DI defined in ζ.
- Recursive Case: For R-Chain(δ,ζ) and R-Chain(ϵ,ζ), R-Chain(η,ζ) is:
- AntecedentNodes(η) $=_{def}$ AntecedentNodes(δ) $\cup$ AntecedentNodes(ϵ).
- ConclusionNode(η) $=_{def}$ ConclusionNode(ϵ).
- AArcs(η) $=_{def}$ AArcs(δ) $\cup$ AArcs(ϵ).
- Condition: ClaimNodes(δ) $\subseteq$ PremissNodes(ϵ).

The condition *connects* one rule to another; it also clarifies the connection between the rules in that δ appears as the "first" rule and ϵ the "second".

An *A(rgument)-Chain* is a subcase of an R-Chain given asserted premisses. The *rationale* is that in logic and philosophy (Hitchcock 2007), an *argument* relates not to just any preferred extension in an AF relative to a rule, but *only* to preferred extensions where the premisses and the conclusion of the rule *both* hold. For SI and DI, this is where the antecedent *and* conclusion are both true. As illustrated below, given a rule and different assertions, different preferred extensions arise, only some of which correlate to the concept of argument. We assume that A-Chains inherit the definitions from R-Chains; i.e. the RuleName of an R-Chain is the ChainName of an A-Chain, the PremiseNodes of a R-Chain are the PremiseNodes of an A-Chain, and so on.

Definition: A-Chains

- Strict A-Chain(η,ζ) if and only if R-Chain(η, ζ) and $\forall\chi \subseteq \zeta_{ANode}$, if χ is a *preferred extension* and AntecedentNodes(η) $\subseteq \chi$, then ConclusionNode(η) $\in \chi$.
- Defeasible A-Chain(η,ζ) if and only if R-Chain(η, ζ) and $\exists\chi \subseteq \zeta_{ANode}$, χ is a *preferred extension*, (AntecedentNodes(η) $\cup$ ConclusionNode(η)) $\subseteq \chi$.

A literal is *justified* by being *asserted* or *by being the conclusion of an A-Chain with justified antecedents*. As with SI and DI, we can assert literals with respect to Strict and Defeasible A-Chains.

Transitivity

Having defined A-Chains, consider how they account for transitivity, which holds of SI but not DI. We consider two cases: SI followed by SI and DI followed by SI. For each, we give relevant cases.

We have in Figure 4 two SIs which form an R-Chain comprised of $P \rightarrow Q$ and $Q \rightarrow R$, where Rule1 is "P strictly implies Q" and Rule3 is "Q strictly implies R".

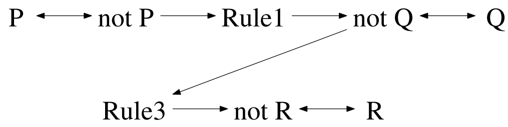

Figure 4: Two SIs

With respect to Figure 4, suppose we look at the one case where we assert *P*. Other results can be calculated along the lines as earlier. This gives the following result:

a. P: {P, Rule1, Q, Rule3, R}.

Figure 4 provides a Strict A-Chain since for every preferred extension where the antecedent *P* holds the conclusion *R* holds; we can say that *R* is *skeptically accepted relative to P*. Though *Q* holds in the extension, it *plays no role* in the inference to R, but is there by default from the SI R-Chain in the first part of the A-Chain. We do not need *Q* to infer that *R* is in the preferred extension, for where *P* is asserted, *Rule1* must hold, from which it follows that *Rule3* and then *R* hold. *Q* is not the antecedent of the A-chain and does not justify that *R* holds. Here, we have two SIs in an A-chain; this demonstrates that closure holds transitively with respect to SI. The *effect* of transitivity emerges from the graph and does not need to be imposed by additional stipulations. We see these two points as important and novel to the literature on the interaction between AFs and logic.

Next consider a combination of DI and SI as in Figure 5 along with two of the results. We have $P \leadsto Q$ where Rule4 is "P defeasibly implies Q" and this is followed by $Q \rightarrow R$ where Rule3 "Q strictly implies R".

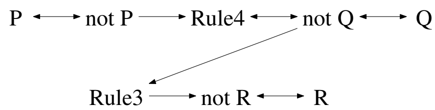

Figure 5: DI followed by SI

a. P: {P, Rule4, Q, Rule3, R}, {P, ¬Q, ¬R}

b. ¬R: {P, ¬Q, ¬R}, {¬P, ¬Q, ¬R}

In the first case, we have a Defeasible A-Chain from *P* to *R*, where *R* is *credulously accepted* relative to *P*. Thus, DI does not support transitivity. In the latter case, we see that *P* is only *credulously accepted* relative to $\neg R$ since contraposition does not hold of defeasible implication. This case is important to our considerations of a problem later.

We see that in virtue of the *structure* we have assigned, we have an AF which represents the difference between SI and DI with respect to transitivity.

Problem Cases

(Caminada & Amgoud 2007) present three problems which, they claimed, any theory of argumentation ought to account for; of these, we discuss the first two for the third is a variant of the second. They provide a *knowledge base* comprised of assertions (strict implications with empty antecedents) and strict and defeasible implications. *Arguments* are constructed from components of the knowledge base or other arguments, which are the sorts of structures we criticise earlier.

Problem 1 – The Married Bachelor

- Strict implications: $\rightarrow$ WR; $\rightarrow$ GO; B $\rightarrow \neg$HW; M $\rightarrow$ HW
- Defeasible implications: WR $\rightsquigarrow$ M; GO $\rightsquigarrow$ B;
- Arguments: A1: [$\rightarrow$ WR]; A2: [$\rightarrow$ GO]; A3: [A1 $\rightsquigarrow$ M]; A4: [A2 $\rightsquigarrow$ B]; A5: [A3 $\rightarrow$ HW]; A6: [A4 $\rightarrow \neg$HW]
- WR means *John wears something that looks like a wedding ring*, M means *John is married*, HW means *Has a wife*, GO means *John often goes out until late with friends*, B means *John is a bachelor*

(Caminada & Amgoud 2007) claim that given this knowledge base and these arguments, both *M* and *B* are incorrectly justified.

In contrast, we represent the knowledge base directly in the AF, with respect to which we assert *GO* and *M*, then reason with respect to A-Chains. This appears in Figure 6. The rule names can be inferred from the structure of the graph.

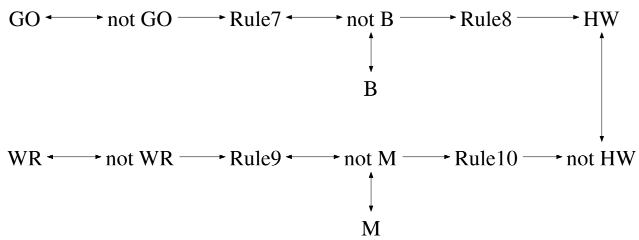

Figure 6: The Married Bachelor Example

With respect to Figure 6, the grounded extension is $\{GO, WR\}$. We have several preferred extensions since *B*, *M*, *HW*, and $\neg HW$ are defeasible. Following our reasoning about SI: if *HW* is in the extension, then *M* and $\neg B$ are in the extension; if $\neg HW$ is in the extension, then *B* and $\neg M$ are in the extension. Neither *M* or *B* are justified in a Strict-A-Chain.

The second problem presented in (Caminada & Amgoud 2007) is formulated as follows:

Problem 2

- Strict implications: $\rightarrow$ A; $\rightarrow$ D; $\rightarrow$ C; B,E $\rightarrow$ Not C
- Defeasible implications: A $\rightsquigarrow$ B; D $\rightsquigarrow$ E;
- Arguments: A1: [[$\rightarrow$ A] $\rightsquigarrow$ B]; A2: [[$\rightarrow$ D] $\rightsquigarrow$ E]; A3: [$\rightarrow$ C]

(Caminada & Amgoud 2007) claim that *A*, *E*, and *C* are justified. They claim that though there is a strict rule which allows us to infer *Not C*, we cannot because it is part of the knowledge base and not the argument network. This, they claim, shows that justified conclusions are not closed under strict rules or could be inconsistent.

We represent the problem as in Figure 7, where we assert *A*, *D*, and *C*. No problem arises: the assertions of *A* and *D* defeasibly imply *B* and *E*. The assertion of *C* implies either $\neg B$ or $\neg E$ strictly holds, which defeats the conjunction of *B* and *E*.

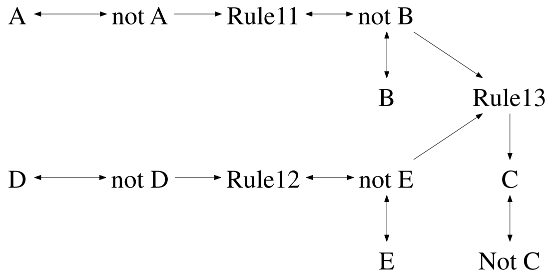

Figure 7: Closure under SI

Discussion

For defeasible rules, the rule and the negation of its consequent can appear in different preferred extensions. We need, then, a systematic means of choosing which one we will accept. Various forms of non-monotonic logic offer us, in effect, systematic ways of making or justifying this choice. One principle, which looks intuitively plausible in our context, would be to choose the preferred extension which maximises the number of rules accepted. This is similar to circumscription (McCarthy 1980), which introduces, for each rule, an additional *ab* predicate which specifies that circumstances are not abnormal with respect to the rule and then works by minimising the extension of the predicate. In our system, we could represent this by adding an *ab* node for every rule, where *ab* attacks the rule. The *ab* node itself attacks and is attacked by a $\neg$*ab* node. We then select the preferred extension with the fewest *ab* nodes. One advantage of including *ab* nodes is that we could now explicitly represent abnormal circumstances: for example *not has wife(X)* would attack the $\neg$*ab* node relating to the rule that a man wearing a gold ring is married.

In this paper, we adapted RBA systems ((Prakken & Sartor 1997), (García & Simari 2004), (Governatori *et al.* 2004), and (Amgoud *et al.* 2004)) into a cohesive and coherent AF, incorporating SI and DI along with a useful and philosophically sound conception of *argument* while retaining the key observations and analyses of AFs. The key novel contribution of the paper is that we have provided a general format to extend the Dungian AFs to reason in the AF relative to an instantiated knowledge base. We have also suggested that the system can be extended further to cover other forms of defeasible reasoning.

References

Amgoud, L.; Caminada, M.; Cayrol, C.; Lagasquie, M.-C.; and Prakken, H. 2004. Towards a consensual formal model: inference part. Technical report, ASPIC project. Deliverable D2.2: Draft Formal Semantics for Inference and Decision-Making.

Bench-Capon, T. J. M. 2003. Persuasion in practical argument using value-based argumentation frameworks. *J. Log. Comput.* 13(3):429–448.

Bondarenko, A.; Dung, P. M.; Kowalski, R. A.; and Toni, F. 1997. An abstract, argumentation-theoretic approach to default reasoning. *Artificial Intelligence* 93:63–101.

Caminada, M., and Amgoud, L. 2007. On the evaluation of argumentation formalisms. *Artif. Intell.* 171(5-6):286–310.

Dung, P. M. 1995. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence* 77(2):321–358.

Dunne, P. E., and Bench-Capon, T. J. M. 2002. Coherence in finite argument systems. *Artificial Intelligence* 141(1):187–203.

García, A. J., and Simari, G. R. 2004. Defeasible logic programming: An argumentative approach. *TPLP* 4(1-2):95–138.

Governatori, G.; Maher, M. J.; Antoniu, G.; and Billington, D. 2004. Argumentation semantics for defeasible logic. *Journal of Logic and Computation* 14(5):675–702.

Hitchcock, D. 2007. Informal logic and the concept of argument. In Gabbay, D.; Thagard, P.; and Woods, J., eds., *Philosophy of Logic (Handbook of the Philosophy of Science)*. Elsevier. 101–130.

McCarthy, J. 1980. Circumscription - a form of nonmonotonic reasoning. *Artificial Intelligence* 13:27–39.

Pollock, J. 1995. *Cognitive Carpentry: A Blueprint for How to Build a Person*. MIT Press.

Prakken, H., and Sartor, G. 1997. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-Classical Logics* 7(1).

Walton, D. 1996. *Argumentation Schemes for Presumptive Reasoning*. Mahwah, N.J.: Erlbaum.

Wyner, A., and Bench-Capon, T. 2008. Modelling judicial context in argumentation frameworks. In Besnard, P.; Doutre, S.; and Hunter, A., eds., *Computational Models of Argument - Proceedings of COMMA 2008*, Frontiers in Artificial Intelligence and Applications. IOS Press. To Appear.

Wyner, A.; Bench-Capon, T.; and Atkinson, K. 2008. Three senses of "argument". In Sartor, G.; Casanovas, P.; Rubino, R.; and Casellas, N., eds., *Computable Models of the Law: Languages, Dialogues, Games, Ontologies*, LNAI 4884, To Appear. Springer.