Relating ways to instantiate abstract argumentation frameworks

Henry Prakken
Department of Information and Computing Sciences, Utrecht University & Faculty of Law, University of Groningen

November 25, 2013

Abstract

This paper studies the relation between various ways to instantiate Dung's abstract argumentation frameworks. First the *ASPIC*$^{+}$ framework, which explicitly generates abstract argumentation frameworks, is equivalently reformulated in terms of John Pollock's recursive labelling method, which does not explicitly generate such frameworks. The reformulation arguably facilitates more natural explanations of dialectical argument evaluation. Then a variant is examined of a recent proposal by Wyner, Bench-Capon and Dunne to instantiate abstract argumentation frameworks without defining explicit inferential relations between arguments. The proposal is reformulated in a way that is equivalent to *ASPIC*$^{+}$ under some limiting assumptions. The proof exploits the equivalence between *ASPIC*$^{+}$ and Pollock's recursive labellings proven in the first part of the paper.

1 Introduction

While most current AI research on argumentation takes its point of departure in [8]'s abstract argumentation framework, there is a tension between approaches which do or do not specify inferential relations between arguments. Examples of the former approach are assumption-based argumentation (initiated in [6]), classical argumentation as studied in [10], and the *ASPIC*($^{+}$) framework of [7, 19, 15] (recently used in [22]). Examples of the latter approach are preference-based argumentation frameworks [1], abstract resolution semantics (initiated in [13]), [4] value-based argumentation frameworks, [2]'s instantiation of these with an argument scheme for practical reasoning, and [23]'s modelling of rule-based argumentation with Dung's frameworks.

As can be seen from these references, Trevor Bench-Capon has throughout his career explored all these ways to use Dung's work. In [4] he took the fully abstract approach, in [2, 23] he defined the structure of arguments but did not allow for inferential relations between them, and in [22] he employed a framework that allows for such relations. It therefore seems appropriate for this volume to further investigate the various ways in which Dung's work can be related to accounts of the structure of argumentation.

In a first contribution I will show that systems for structured argumentation do not have to explicitly produce a Dung-style abstract argumentation framework. I will do so by proving an equivalence between *ASPIC*$^+$ as formulated in [19, 15] (which explicitly produces Dung-style abstract argumentation frameworks) and a reformulation of *ASPIC*$^+$ in terms of John Pollock's [17] recursive labelling method (which does not follow this approach). This arguably facilitates more natural explanations of the outcome of dialectical argument evaluation. In a second contribution I will investigate [23]'s way to model structured argumentation without defining explicit inferential relations between arguments. It will turn out that this approach can be formulated in a way that is equivalent to *ASPIC*$^+$ if some limiting assumptions are made. The proof exploits the equivalence between *ASPIC*$^+$ and Pollock's recursive labellings proven in the first part of this paper. I will then conclude with some observations on the pros and cons of abstract models of argumentation.

2 Basic formalisms

In this section I review [8]'s abstract argumentation frameworks and the *ASPIC*$^+$ framework. An *abstract argumentation framework* (AF) is a pair $(\mathcal{A}, \mathcal{D})$, where $\mathcal{A}$ is a set of *arguments* and $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$ is a binary relation of *defeat*. A semantics for *AFs* returns sets of arguments called *extensions*, which are internally conflict-free and defend themselves against attack. One way to characterise the various semantics is with *labellings*.

Definition 1 A *labelling* of an abstract argumentation framework $(\mathcal{A}, \mathcal{D})$ is any assignment of either the label *in* or *out* (but not both) to zero or more arguments from $\mathcal{A}$ such that:

1. an argument is *in* iff all arguments defeating it are *out*.
2. an argument is *out* iff it is defeated by an argument that is *in*.

Then *stable semantics* labels all arguments, while *grounded semantics* minimises and *preferred semantics* maximises the set of arguments that are labelled *in* and *complete semantics* allows any labelling. Relative to a semantics, an argument is *justified* on the basis of an AF if it is labelled *in* in all labellings, it is *overruled* if it is labelled *out* in all labellings, and it is *defensible* if it is neither justified nor overruled.

The *ASPIC*$^+$ framework [19] gives structure to Dung's arguments and defeat relation. It defines arguments as inference trees formed by applying strict or defeasible inference rules to premises formulated in some logical language. Informally, if an inference rule's antecedents are accepted, then if the rule is strict, its consequent must be accepted *no matter what*, while if the rule is defeasible, its consequent must be accepted *if there are no good reasons not to accept it*. Arguments can be attacked on their (non-axiom) premises and on their applications of defeasible inference rules. Some attacks succeed as *defeats*, which is partly determined by preferences. The acceptability status of arguments is then defined by applying any of [8] semantics for abstract argumentation frameworks to the resulting set of arguments with its defeat relation.

ASPIC$^+$ is not a system but a framework for specifying systems. It defines the notion of an abstract *argumentation system* as a structure consisting of a logical language $\mathcal{L}$ closed under negation[1], a set $\mathcal{R}$ consisting of two subsets $\mathcal{R}_s$ and $\mathcal{R}_d$ of strict and defeasible inference rules, and a naming convention n in $\mathcal{L}$ for defeasible rules in order to talk about the applicability of defeasible rules in $\mathcal{L}$. Thus, informally, $n(r)$ is a wff in $\mathcal{L}$ which says that rule $r \in \mathcal{R}$ is applicable. *ASPIC*$^+$ as a framework does not make any assumptions on how the elements of an argumentation system are defined. In *ASPIC*$^+$ argumentation systems are applied to knowledge bases to generate arguments and counterarguments. Combining these with an argument ordering results in argumentation theories, which generate Dung-style AFs.

Definition 2 [**Argumentation systems**] An *argumentation system* is a triple $AS = (\mathcal{L}, \mathcal{R}, n)$ where:

- $\mathcal{L}$ is a logical language closed under negation ($\neg$).
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ is a set of strict ($\mathcal{R}_s$) and defeasible ($\mathcal{R}_d$) inference rules of the form $\varphi_1, \ldots, \varphi_n \to \varphi$ and $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ respectively (where φ_i, φ are meta-variables ranging over wff in $\mathcal{L}$), and $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- $n : \mathcal{R}_d \longrightarrow \mathcal{L}$ is a naming convention for defeasible rules.

We write $\psi = -\varphi$ just in case $\psi = \neg\varphi$ or $\varphi = \neg\psi$.

Definition 3 [**Knowledge bases**] A *knowledge base* in an $AS = (\mathcal{L}, \mathcal{R}, n)$ is a set $\mathcal{K} \subseteq \mathcal{L}$ consisting of two disjoint subsets $\mathcal{K}_n$ (the *axioms*) and $\mathcal{K}_p$ (the *ordinary premises*).

Intuitively, the axioms are certain knowledge and thus cannot be attacked, whereas the ordinary premises are uncertain and thus can be attacked.

Arguments can be constructed step-by-step from knowledge bases by chaining inference rules into trees. In what follows, for a given argument the function `Prem` returns all its premises, `Conc` returns its conclusion and `Sub` returns all its sub-arguments.

Definition 4 [**Arguments**] An *argument* A on the basis of a knowledge base KB in an argumentation system $(\mathcal{L}, \mathcal{R}, n)$ is:

1. φ if $\varphi \in \mathcal{K}$ with: $\texttt{Prem}(A) = \{\varphi\}$; $\texttt{Conc}(A) = \varphi$; $\texttt{Sub}(A) = \{\varphi\}$; $\texttt{TopRule}(A)$ = undefined.
2. $A_1, \ldots A_n \to/\Rightarrow \psi$ if $A_1, \ldots, A_n$ are arguments such that there exists a strict/defeasible rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \to/\Rightarrow \psi$ in $\mathcal{R}_s/\mathcal{R}_d$.
   $\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$, $\texttt{Conc}(A) = \psi$, $\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$; $\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \to/\Rightarrow \psi$.

Arguments can be attacked in three ways: on their premises (undermining attack), on their conclusion (rebutting attack) or on an inference step (undercutting attack). The latter two are only possible on applications of defeasible inference rules.

[1] In most papers on *ASPIC*$^+$ negation can be non-symmetric, an idea taken from [6]. In this paper we present the special case with symmetric negation.

Definition 5 [**Attack**] A *attacks* B iff A *undercuts*, *rebuts* or *undermines* B, where:

• A *undercuts* argument B (on B') iff $\texttt{Conc}(A) = -n(r)$ and $B' \in \texttt{Sub}(B)$ such that B''s top rule r is defeasible.

• A *rebuts* argument B (on B') iff $\texttt{Conc}(A) = -\varphi$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$.

• Argument A *undermines* B (on B') iff $\texttt{Conc}(A) = -\varphi$ for some $B' = \varphi$, $\varphi \notin \mathcal{K}_n$.

Undercutting attacks succeed as *defeats* independently of preferences over arguments, since they express exceptions to defeasible inference rules. Rebutting and undermining attacks succeed only if the attacked argument is not stronger than the attacking argument.

Definition 6 [**Defeat**] A *defeats* B iff:A undercuts B, or; A rebuts/undermines B on B' and $A \nprec B'$.

Note that the definitions of attack and defeat explicitly take the subargument relations between arguments into account. In fact, it was recently shown in [20] and [14] that instantiations of Dung's framework that do not do so are severely limited in applicability and run the risk of producing counterintuitive consequences.

Definition 7 Let AT be an *argumentation theory* (AS, KB). A *structured argumentation framework (*SAF*)* defined by AT, is a triple $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ where $\mathcal{A}$ is the set of all finite arguments constructed from KB in AS, $\preceq$ is an ordering on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff X attacks Y.

Abstract argumentation frameworks are then generated as follows:

Definition 8 [**Argumentation frameworks**] An *abstract argumentation framework (AF) corresponding to a* SAF = $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ is a pair $(\mathcal{A}, D)$ such that D is the defeat relation on $\mathcal{A}$ determined by SAF.

Now one way to define the justification status of an argument is as follows (other definitions are possible but for present purposes they do not matter):

Definition 9 A wff $\varphi \in \mathcal{L}$ is *justified* if φ is the conclusion of a justified argument, *defensible* if φ is not justified but the conclusion of a defensible argument, and *overruled* if φ is defeated by a justified argument.

3 Labellings vs. recursive labellings

My view on [8]'s work has always been that it has to be used as the final component of a full-fledged argumentation logic, to define a nonmonotonic consequence notion for the logic. First the structure of arguments must be defined, which will in general be recursive to reflect the step-by-step nature of argumentation. Then a defeat relation must be defined, which will in general have to distinguish between direct and indirect defeat, because of the recursive structure of arguments. Then the resulting set of arguments and the defeat relation defined over it are taken to be a Dung AF. Adopting a semantics for the AF then yields the consequence notion for the logic.

While theoretically this works fine, a practical drawback is that in the thus induced Dung *AF* the distinction between direct and indirect defeat is left implicit, since in a Dung *AF* the subargument relation between arguments cannot be shown. For presentation purposes this is somewhat unnatural. Consider the following example of a civil legal case, adapted from [21]. Assume that in a medical malpractice case, a doctor is liable for compensation if the patient was injured because of the doctor's negligence, and that if a patient is injured in a non-risky operation, this is negligence. We also have that an appendicitis operation generally is a non-risky operation but that operations on patients with bad blood circulation are generally risky. Assume finally, that a given patient was injured in an appendicitis operation and that two medical tests gave contradicting results on whether the patient had bad blood circulation. This all can be represented with the following facts and defeasible rules ($\mathcal{R}_d$ consists of r_1-r_6 while $\mathcal{K}$ consists of f_1-f_4; note that for convenience we here use the *ASPIC*$^+$ rules in domain-specific ways).

r_1:	*injury*, *negligence* $\Rightarrow$ *compensation*	f_1:	*injury*
r_2:	*injury*, $\neg$ *risky operation* $\Rightarrow$ *negligence*	f_2:	*appendicitis*
r_3:	*appendicitis* $\Rightarrow \neg$ *riskyOperation*	f_3:	*medicalTest1*
r_4:	*badCirculation* $\Rightarrow$ *riskyOperation*	f_4:	*medicalTest2*
r_5:	*medicalTest1* $\Rightarrow$ *badCirculation*		
r_6:	*medicalTest2* $\Rightarrow \neg$ *badCirculation*		

We then have the following arguments:

A_1:	*injury*	B_1:	*medicalTests1*
A_2:	*appendicitis*	B_2:	$B_1 \Rightarrow$ *badCirculation*
A_3:	$A_2 \Rightarrow \neg$ *riskyOperation*	B_3:	$B_2 \Rightarrow$ *riskyOperation*
A_4:	$A_1, A_3 \Rightarrow$ *negligence*		
A_5:	$A_1, A_4 \Rightarrow$ *compensation*	C_1:	*medicalTests2*
		C_2:	$C_1 \Rightarrow \neg$ *badCirculation*

Let us first concentrate on the attack relations between the arguments. At first sight it would seem that we have three arguments $A = A_1$-A_5, $B = B_1$-B_3 and $C = C_1$-C_2, such that both A and B and B and C mutually attack each other as in Figure 1. However, upon closer investigation things are more complicated. First, we have not

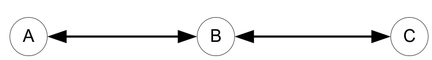

Figure 1: Abstract attack graph (simplistic)

three but ten arguments. Second, we have not four but seven attack relations: A_3 and B_3 directly rebut each other and as a consequence B_3 indirectly rebuts both A_4 and A_5, namely on A_3. Likewise, B_2 and C_2 directly rebut each other and as a consequence C_2 indirectly rebuts B_3, namely on B_2. So the full abstract attack graph induced by the *ASPIC*$^+$ attack relations is as in Figure 2. In this graph the information is lost that some attack arrows duplicate other attack arrows. For example, the arrows from B_3 to A_4 and A_5 duplicate the arrow from B_3 to A_3.

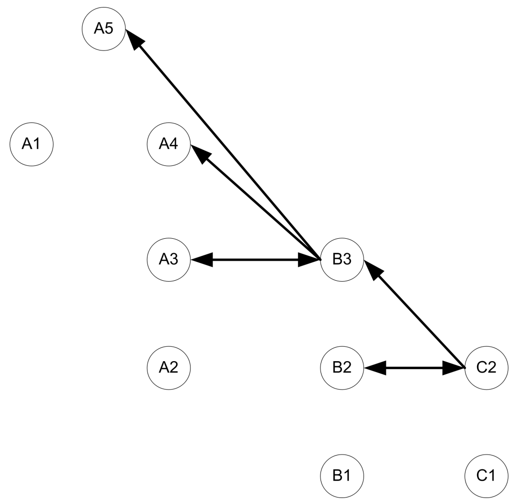

Figure 2: Abstract attack graph (correct)

Therefore, this graph does not adequately show a decision maker which conflicts have to be resolved. When faced with this graph, a decision maker might think that s/he has to determine of seven attack relations whether they result in defeat, but in fact s/he has to determine this for only four attack relations, namely, those between A_3 and B_2 and between B_2 and C_2. (This in fact illustrates the problems with Preference-based abstract *AFs* [1] and abstract resolution semantics [3], since these approaches cannot recognise that some defeat relations duplicate other defeat relations.) If the attacks are resolved according to Definition 6 then the indirectly defeat relations are correctly computed on the basis of the direct defeat relations, resulting in an abstract argumentation framework with the intuitive outcome. For example, suppose that $A_3 \prec B_3$ and $B_2 \prec C_2$. Then we obtain Figure 3 (which shows the unique labelling in all semantics, in which the arguments that are *in* are coloured gray). However, even though this outcome is intuitive, the process of obtaining it from the abstract attack graph is not well explained by these graphics.

Moreover, the result of the abstract *AF* has to be transported back with Definition 9 to the *SAF* that induced it, which tells us that B_2 and B_3 are *out* in all labellings while all other arguments are *in* in all labellings. Graphically this can be displayed as in Figure 4. Now this graph suggests a more intuitive way to explain how the status of A_5 can be evaluated: A_5 is *in* since it has no direct defeaters and both of its direct subarguments A_1 and A_4 are *in*. In turn, A_1 is in since it is an undefeated fact while A_4 is in since it has no direct defeaters and both of its direct subarguments A_1 and A_3 are *in*. In turn, A_3 is *in* since, firstly, its direct defeater B_3 is *out* and, secondly, its direct subargument A_2 is *in*since it is an undefeated fact. Finally, B_3 is *out* since it has a direct defeater that is *in* namely, C_2, which is is preferred over B_3 and which direct subargument C_3 is *in* since it is an undefeated fact.

Such an explanation arguably shows more clearly why an argument has a certain

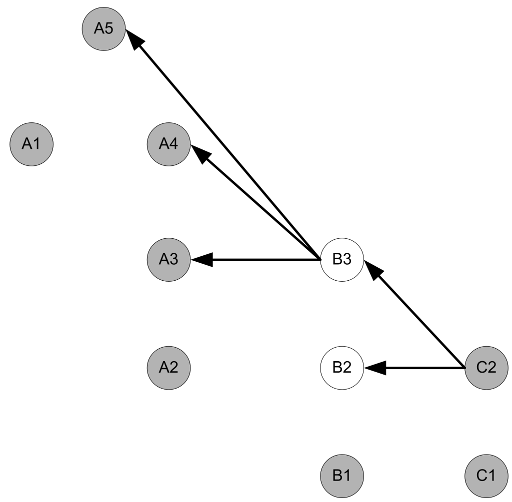

Figure 3: Abstract argumentation framework

status, since it looks at both the defeat relations and the inferential support relations between the arguments. In fact, such an explanation directly follows Pollock's [17] recursive definition of 'defeat status assignments' (two other systems with similar recursive definitions are Defeasible Logic [16] and Carneades [9]). I will therefore in the next section adapt this definition to *ASPIC*$^+$ and then prove equivalence with Definition 1 of labellings for *ASPIC*$^+$.

4 Recursive labellings

To adapt Pollock's [17] recursive definition of 'defeat status assignments' to *ASPIC*$^+$, first the definitions of attack and defeat must be adapted; as in [17] they now only capture direct attack and defeat while their indirect counterparts will now be considered by the recursive labelling definition.

Definition 10 [**p-Attack**] A *p-attacks* B iff A *p-undercuts*, *p-rebuts* or *p-undermines* B, where:

- A *p-undercuts* argument B iff $\texttt{Conc}(A) = -n(r)$ and B has a defeasible top rule r.
- A *p-rebuts* argument B iff $\texttt{Conc}(A) = -\texttt{Conc}(B)$ and B has a defeasible top rule.
- Argument A *p-undermines* B iff $\texttt{Conc}(A) = -\varphi$ and $B = \varphi$, $\varphi \notin \mathcal{K}_n$.

In our example we have that A_3 and B_3 p-attack each other but B_2 does not p-attack A_4 or A_5. Likewise, B_2 and C_2 p-attack each other but C_2 does not p-attack B_3.

Definition 11 [**p-Defeat**] A *p-defeats* B iff: A p-undercuts B, or; A p-rebuts/p-undermines B and $A \nprec B$.

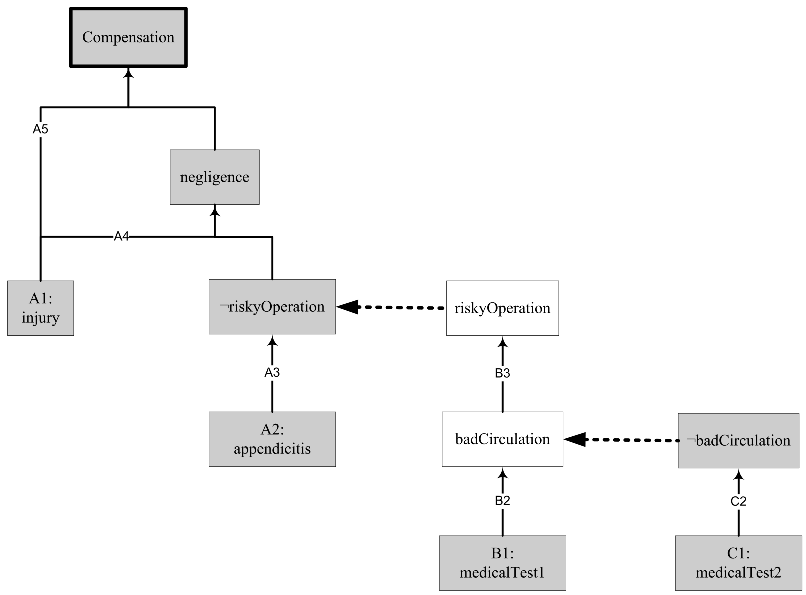

Figure 4: ASPIC+ framework

It is easy to verify that:

Proposition 1 A defeats B iff A p-defeats B or a proper subargument of B.

Proof:

- A undermines B iff:
  $\texttt{Conc}(A) = -\varphi$ and $B = \varphi$, $\varphi \notin \mathcal{K}_n$, or A undermines some subargument B' of B such that B' is a premise of B; iff:
  A p-defeats B or a proper subargument of B.

- A rebuts B iff:
  $\texttt{Conc}(A) = -\texttt{Conc}(B)$ and B has a defeasible top rule or for some subargument B' of B it holds that $\texttt{Conc}(A) = -\texttt{Conc}(B')$ and B' has a defeasible top rule; iff:
  A p-rebuts B or a proper subargument of B.

- A undercuts B iff:
  $\texttt{Conc}(A) = -n(r)$ and B has a defeasible top rule r or for some subargument B' of B it holds that $\texttt{Conc}(A) = -n(r)$ and B' has a defeasible top rule r; iff:
  A p-undercuts B or a proper subargument of B.

Then the proposition easily follows. □

Definition 12 A *p-structured argumentation framework (pSAF)* defined by an argumentation theory $AT = (AS, KB)$ is a triple $\langle \mathcal{A}, \mathcal{C}^p, \preceq \rangle$ where $\mathcal{A}$ is the set of all arguments constructed from KB in AS, $\preceq$ is an ordering on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}^p$ iff X p-attacks Y.

Definition 13 [**p-Argumentation frameworks**] A *p-abstract argumentation framework (pAF) corresponding to a pSAF* = $\langle \mathcal{A}, \mathcal{C}^p, \preceq \rangle$ is a pair $(\mathcal{A}, D)$ such that D is the p-defeat relation on $\mathcal{A}$ determined by *pSAF*.

Definition 14 [p-labellings.]

1. (In, Out) is a *p-labelling* of a *pAF* iff $In \cap Out = \emptyset$ and for all $A \in \mathcal{A}_{pAF}$ it holds that:
   - (a) A is labelled *in* iff:
     - i. All arguments that p-defeat A are labelled *out*; and
     - ii. If A is of the form $B_1, \ldots, B_n \rightarrow / \Rightarrow \varphi$ then all of $B_1, \ldots, B_n$ are labelled *in*; and
   - (b) A is labelled *out* iff:
     - i. A is p-defeated by an argument that is labelled *in*; or
     - ii. A is of the form $B_1, \ldots, B_n \rightarrow / \Rightarrow \varphi$ and some of $B_1, \ldots, B_n$ are labelled *out*.

   The notions of complete, stable, preferred and grounded labellings are defined as above.

Theorem 2 (In, Out) *is a p-labelling of pAF corresponding to pSAF* $= \langle \mathcal{A}, \mathcal{C}^p, \preceq \rangle$ *iff* (In, Out) *is a labelling of AF corresponding to SAF* $= \langle \mathcal{A}, \mathcal{C}, \preceq \rangle$.

Proof:
From left to right, assume first that A is *in*. We first prove by induction on the structure of A that all subarguments of A are *in*. For $A \in \mathcal{K}$ this holds trivially. If A is of the form $B_1, \ldots, B_n \rightarrow / \Rightarrow \varphi$ then if A is *in*, by definition of p-labellings also all of $B_1, \ldots, B_n$ are *in*. By the induction hypothesis all subarguments of all of $B_1, \ldots, B_n$ are also *in*, but then all subarguments of A are *in*. Now since A is *in*, we have that all B that p-defeat A are *out* and all subarguments of A are *in*. Then all B that p-defeat A are *out* and all B that p-defeat a subargument of A are *out*, but then all B that defeat A are *out*. So constraint (1) on labellings is satisfied.

Assume next that A is *out*. We prove by induction that A is defeated by some argument that is *in*, in which case constraint (2) on labellings is satisfied. For $A \in \mathcal{K}$ this holds trivially, while if A is of the form $B_1, \ldots, B_n \rightarrow / \Rightarrow \varphi$ then either some B that p-defeats A is *in* or some B_i is *out*. In the first case, B also defeats A. In the second case, by the induction hypothesis B_i is defeated by an argument that is *in*. But then A is defeated by the same argument.

From right to left, assume first that A is *in*. Then all B that defeat A are *out*. But then all B that p-defeat A or a proper subargument of A are *out*. Next we prove by induction that all proper subarguments of A are *in*. This is trivial if $A \in \mathcal{K}$. Suppose next A is of the form $B_1, \ldots, B_n \rightarrow / \Rightarrow \varphi$. Then by the induction hypothesis for all of $B_1, \ldots, B_n$, all their proper subarguments, including the immediate ones, are in. Moreover, by assumption all p-defeaters of any such B_i are *out*. But then all B_i are *in*. Then constraint (1) on p-labellings is satisfied.

Assume next that A is *out*. Then A is defeated by an argument that is *in*. But then either A or a proper subargument A' of A is p-defeated by an argument B that is *in*. In the first case, constraint (2) of p-labellings is satisfied. In the second case A is of the form $B_1, \ldots, B_n \rightarrow / \Rightarrow \varphi$. Note that A' is *out* and that A' is a subargument of some B_i. But then B defeats B_i, so since B is *in* we have that B_i is *out*. So constraint (2) on p-labellings is satisfied. □

Since the sets that are labelled, respectively, p-labelled are the same, it immediately follows that the equivalence also holds for the notions of stable, preferred and grounded labellings.

This result is very similar to Hadassa Jakobovits' results in [11, 12] that Pollock's [17] 'labellings' for his inference graphs are equivalent to preferred semantics and that his 'partial labellings' are equivalent to her 'rooted complete labellings'. The main difference is that Pollock's inference graphs are not the same as *ASPIC*$^+$ SAFs (although equivalence should be easy to prove).

With these equivalences established, there is no need any more to translate an *ASPIC*$^+$ argumentation theory into a Dung-style abstract argumentation framework. Instead, the labellings of arguments and the corresponding status of their conclusions can be shown inductively, which, as explained at the end of Section 3, is arguably more natural.

5 Argument structure in purely abstract frameworks

In the second part of the paper I want investigate an idea originally proposed by [23]. Their idea is that arguments in a Dung AF are instantiated as a rule and that the step-by-step nature of arguments is induced by allowing such rules to be attacked by special arguments challenging one of their antecedents, and to in turn allow these arguments to be attacked by a rule with as consequent the challenged antecedent. I will formalise this idea in the definition of a wbd-argumentation theory (*wbd-AF*), in which, given an *ASPIC*$^+$ argumentation theory, arguments are either rules or items from the knowledge base or *why* φ moves for any φ that is an antecedent of some rule. My aim then is to translate such a *wbd-AF* into an *ASPIC*$^+$-induced *pAF* and then to investigate correspondences between the labellings of a *wbd-AF* and the p-labellings of the corresponding *pAF* as defined above. If any correspondence is found, then the results of the previous section make it carry over to the Dungean semantics of *ASPIC*$^+$-induced *AFs*.

It will turn out that correspondences can be found but only under some limiting assumptions. The point is that in the new approach the success of an attack as defeat should completely depend on the last of the arguments, that is, on their top rules. This at least requires that arguments are compared with a last-link argument ordering. Moreover, it excludes the use of non-axiom premises and even the use of strict rules. To see the latter, consider the following example, in which a defeasible rule is attacked by a strict rule:

$p \Rightarrow_{r1} q$
$r \rightarrow \neg q$

In *ASPIC*$^+$ with last-link comparison, whether a rebuttal with a strict top rule succeeds depends on how the antecedents of the strict rule (in this case r) are derived. For example, if we have

$s \Rightarrow_{r2} r$
$r_1 < r_2$

Then the attack with the strict rule succeeds but if we have

$t \Rightarrow_{r3} r$
$r_r < r_1$

then the attack with the strict rule does not succeed. So whether the attack with the strict rule succeeds depends on how a *why* r attack on the strict rule is counterattacked. So to make [23]'s approach work in general the argument ordering of the wbd-style arguments must be made dependent on the ways the antecedents of the rules can be derived but this defeats the idea of the approach (this is why the nodes in [17]'s defeat graphs are not just well-formed formulas but 'lines of argument' encoding the way a formula is derived; thus if a formula can be derived in more than one way, the graph contains more than one node with the same formula). In fact, this is a problem for any model that defines the evaluation of arguments in a top-down manner without fully constructing the argument before evaluating it, such as also my own dialogue framework in [18]. In that paper I made an assumption that 'backwards extending' an argument in reply to *why* moves cannot make an argument weaker, but then I did not realise how strong that assumption is.

For these reasons I will from now on assume that the set of strict rules is empty and that all premises are axioms. Then with the last link ordering the preference relation between two rule arguments solely depends on their last rules. Note also that in *ASPIC*$^+$ rebutting an argument with an item from $\mathcal{K}_n$ always succeeds, since strict-and-firm arguments (that is, arguments with only strict rules and only axiom premises) are always preferred over all other arguments.

As for the attack relations in the wbd approach, defeasible rules are attacked by any undercutting rule, by any defeasible rule with contradictory consequent, by any item from the knowledge base that contradicts its consequent, and by a *why* φ move for any antecedent φ of the rule. Items from the knowledge base have no attackers while *why* φ moves are attacked by any rule with consequent φ. Table 1 lists all these well-formed arguments and attacks (it assumes all mentioned rules to be in $\mathcal{R}_d$; for convenience, their informal name is sometimes subscripted to the arrow). The defeat relation now equals the attack relation except that an attack of a defeasible rule with a contradictory consequent only succeeds if the attacker is not inferior to its target, that is, only if $r' \not< r$. This induces a *wbd-AF*, that is, a set of arguments in the just-defined sense with a binary relation of defeat. Formally all this amounts to the following:

Definition 15 Let AS be any argumentation system with no strict rules, KB a knowledge base with no ordinary premises, A^g the set of all arguments given KB and AS as

Table 1: wbd arguments and attacks

Arguments	Attacks
$\varphi\ (\varphi \in \mathcal{K})$	
$\varphi_1, \ldots, \varphi_n \Rightarrow_r \varphi$	*why* $\varphi_i\ (\varphi_i \in \{\varphi_1, \ldots, \varphi_n\})$ $-\varphi\ (\varphi \in \mathcal{K})$ $-n(r)\ (\varphi \in \mathcal{K})$ $\varphi_1, \ldots, \varphi_n \Rightarrow_{r'} -\varphi$
why φ	$\varphi_1, \ldots, \varphi_n \Rightarrow_{r'} -n(r)$ $\varphi_1, \ldots, \varphi_n \Rightarrow_r \varphi$ $\varphi\ (\varphi \in \mathcal{K})$

defined in Table 1, $\mathcal{C}^g$ the attack relation on A^g as defined in Table 1 and $\leq$ an ordering on $\mathcal{R}_d$. Then the *wbd argumentation framework* corresponding to $(AS, KB, \leq)$ is a pair *wbd-AF* $= (A^g, \mathcal{D}^g)$ where $\mathcal{D}^g$ is the defeat relation on $\mathcal{A}^g$ defined by $\mathcal{C}^g$ and $\leq$ as follows: A defeats B iff A attacks B and, in case the attack is as in the fifth line of Table 1, $r' \not< r$.

A *pSAF corresponding to* $(AS, KB, \leq)$ is a triple $(\mathcal{A}, \mathcal{C}, \preceq)$ where $\preceq$ is the last-link ordering on $\mathcal{A}$ induced by $\leq$ as defined in [15]. A *pAF corresponding to* $(AS, KB, \leq)$ is the *pAF* corresponding to *pSAF* as defined in Definition 13.

A *wbd-AF* is *completed* if it (1) contains every possible attack and (2) has no unanswered *why* attacks and (3) $\mathcal{K}$ and $\mathcal{R}_d$ are defined such that no attack cycles through *why* moves are possible (that is, circular arguments are excluded).

Labellings of completed *wbd-AFs* are now defined in the usual way of Definition 1.

In fact, I am only interested in *wbd-AFs* that are 'grounded' in the knowledge base. Therefore, from now on I will implicitly assume that all *wbd-AFs* are completed. Then a given *wbd-AF* is converted to a *pAF* as follows:

Definition 16 For any *wbd-AF* $G = (\mathcal{A}^g, \mathcal{D}^g)$ the corresponding *pAF* $G^p = (\mathcal{A}^p, \mathcal{D}^p)$ is inductively defined as follows.

1. $A \in \mathcal{A}^p$ iff:
   (a) $A \in \mathcal{A}^g$ and $A \in \mathcal{K}$; or
   (b) A is of the form $B_1, \ldots, B_n \Rightarrow \varphi$ and $\texttt{Conc}(B_1), \ldots, \texttt{Conc}(B_n) \Rightarrow \varphi \in \mathcal{A}^g$ and all of $B_1, \ldots, B_n$ are in $\mathcal{A}^p$.
2. $(A, B) \in \mathcal{D}^p$ iff $A, B \in \mathcal{A}^p$ and B has a top rule r and (i) $A \in \mathcal{K}$ or (ii) A has a top rule $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow_r' \varphi$ and $(r', r) \in \mathcal{D}^g$.

Figure 5 visualises the *wbd-AF* corresponding to the example of Section 3. It assumes that $r_5 < r_6$ and $r_3 < r_4$, corresponding to the last-link argument ordering $A_3 \prec B_3$ and $B_2 \prec C_2$ of our medical negligence example[2]

[2]Defeasible rules are written with $\rightarrow$ instead of $\Rightarrow$ since my drawing program has no double arrow.

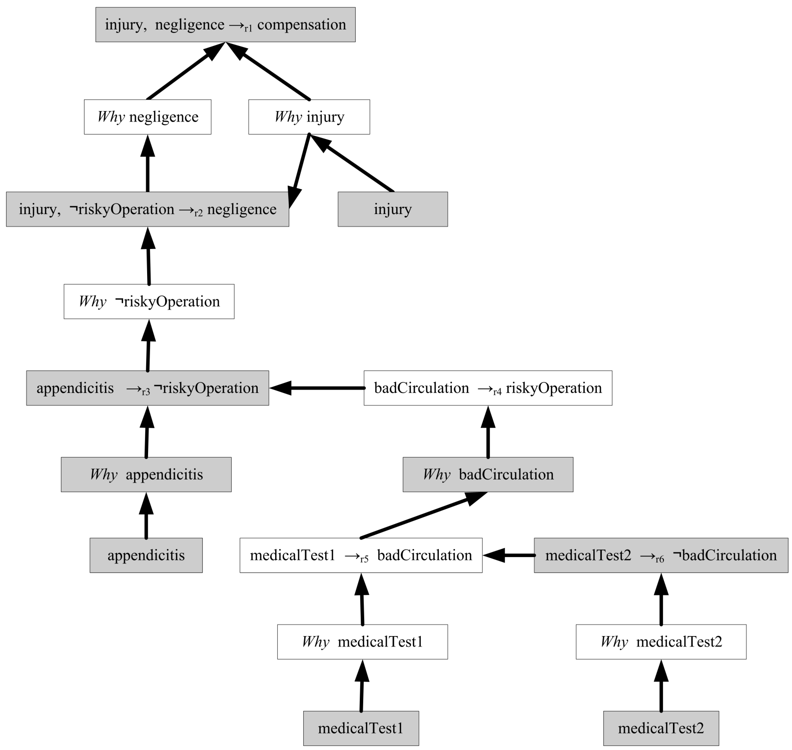

Figure 5: wbd argumentation framework

Proposition 3 For any *wbd-AF* $G = (\mathcal{A}^g, \mathcal{D}^g)$ corresponding to $(KB, AS, \leq)$ with corresponding *pAF* $G^p = (\mathcal{A}^p, \mathcal{D}^p)$, and any *pSAF* $= (\mathcal{A}, \mathcal{D})$:

1. A is an argument in $\mathcal{A}^p$ iff A is an argument in $\mathcal{A}$.
2. $(A, B) \in \mathcal{D}^p$ iff $(A, B) \in \mathcal{D}$.

Proof:
Proof of (1): From left to right follows from the fact that all arguments in $\mathcal{A}^p$ only use elements from $\mathcal{R}_d$ and $\mathcal{K}$ and the structural similarity between Definition 16(1,2) and Definition 4. From right to left is proven with induction on the structure of arguments. For elements of $\mathcal{K}$ this follows immediately from the fact that G is completed. Consider next any argument on the basis of AT of the form $A_1, \ldots, A_n \Rightarrow \varphi$. Then by the induction hypothesis all of $A_1, \ldots, A_n$ are in $\mathcal{A}^p$. Moreover, by construction of $\mathcal{A}^p$ we have that for all such A_i, either $A_i \in \mathcal{K}$ so $A_i \in \mathcal{A}^g$ or A_i is of the form $\texttt{Conc}(B_1), \ldots, \texttt{Conc}(B_n) \Rightarrow_{ri} \varphi$ so $r_i \in \mathcal{A}^g$. But then any *why* $\texttt{Conc}(A_i)$ attack on A can be counterattacked by A_i or r_i, so since G is completed, $A \in \mathcal{A}^g$. But then $A \in \mathcal{A}^p$ by clause (2) of Definition 16.
Proof of (2):

From left to right: If $(A, B) \in \mathcal{D}^p$ then B has a defeasible top rule r. If $A \in \mathcal{K}$ or

A has conclusion $-n(r)$ then A p-defeats B regardless of the rule priorities. If A has a defeasible top rule r' then $r' \not< r$ since $(r', r) \in \mathcal{D}^g$. But then A successfully p-rebuts B, so A p-defeats B.

From right to left, note first that by (1) both A and B are in $\mathcal{A}^p$ and then by definition of $\mathcal{A}^p$, A and B are either themselves in $\mathcal{A}^g$ as elements of $\mathcal{K}$ or their top rules are in $\mathcal{A}^g$. Then because of the restrictive assumptions on AT in Definition 15 it holds that A p-defeats B iff B has a defeasible top rule r and either (1) $A \in \mathcal{K}$, in which case $(A, B) \in \mathcal{D}^p$ since $(A, r) \in \mathcal{D}^g$ by row (3) of Table 1, or (2) A has a defeasible top rule r' with (2a) consequent $-n(r)$, in which case $(A, B) \in \mathcal{D}^p$ since $(r', r) \in \mathcal{D}^g$ by row (4) of Table 1, or with (2b) consequent $-\varphi$ (where φ is the consequent of r), in which case $(A, B) \in \mathcal{D}^p$ since $(r', r) \in \mathcal{D}^g$ by row (5) of Table 1 and $r' \not< r$. $\square$

It can be shown that in a *wbd-AF* all arguments in the corresponding *pAF* are 'hidden'. Such hidden arguments are first defined as follows, after which it will be shown that together they correspond to the arguments of the corresponding *pAF*.

Definition 17 For each non-why argument A in G a directed acyclic graph $T = (N, L)$ is inductively defined as follows:

1. $(N_0, L_0) = (\{A\}, \emptyset)$
2.
   (a) $N_{i+1} = N_i \cup \{A^*\}$ if A^* is some argument in G that attacks some why-attacker B of some argument $A^i \in N_i$ such that no attacker of B is in N_i, otherwise $N_{i+n} = N_i$;

   (b) $L_{i+1} = L_i \cup \{(A^*, A^i) \mid A^i \in N_i \text{ and } \texttt{Conc}(A^*) \in \texttt{Ant}(A^i)\}$ if some A^* as defined under (a) exists, otherwise $L_{i+1} = L_i$.

T^A denotes for any non-why argument $A \in G$ the set of all such graphs T for A, while T^G denotes the set of all A^T for any non-why argument A in G.

Definition 18 Let G be a *wbd-AF* with corresponding *pAF* G^p. For any $A^T \in T^G$, the corresponding argument A^p is defined as follows:

1. If $A^T \in \mathcal{K}$ then $A^p = A^T$;
2. If the root of A^T is in $\mathcal{R}^d$ then A^p is of the form $A_1, \ldots, A_n \Rightarrow \texttt{Conc}(T)$, where $A_1, \ldots, A_n$ correspond to the subtrees $A^{T1}, \ldots, A^{Tn}$ of A^T starting with A's children.

Definitions 17 and 18 are illustrated by the structural similarity of Figures 4 and 5. The idea is that in Figure 5 all *why* moves are removed and the thus disconnected rules and facts are connected by support links.

Proposition 4 For any *wbd-AF* G with corresponding *pAF* G^p:

1. for any $A^T \in T^G$ its corresponding argument A^p is in G^p;

2. for any argument A in G^p there exists a unique $T \in T^G$ such that A corresponds to T.

Proof:
Proof of (1): For elements in $\mathcal{K}$ this is obvious. Next, the induction hypothesis is that for any $A^T \in T^G$ all trees $B^{T1}, \ldots, B^{Tn}$ for all children $B_1, \ldots, B_n$ of A in T correspond to an argument in G^p. By construction of A^T we have that G contains a rule $\texttt{Conc}(B_1), \ldots, \texttt{Conc}(B_n) \Rightarrow \texttt{Conc}(A)$ so by construction of G^p it holds that $B_1, \ldots, B_n \Rightarrow \texttt{Conc}(A)$ is an argument in G^p; moreover, this argument corresponds to A^T.
Proof of (2): Suppose $A \in G^p$. If $A \in \mathcal{K}$ then $(A, \emptyset) \in T^G$. Assume next A is of the form $B_1, \ldots, B_n \Rightarrow B$ and all B_i correspond to some unique $T_i \in T^G$. By construction of G^p we have that $\texttt{Conc}(B_1), \ldots, \texttt{Conc}(B_n) \Rightarrow \texttt{Conc}(A) \in G$. Then since G is completed, $(N, L) \in T^G$ where $N = T_1, \ldots, T_n \cup \{A\}$ and $L = L_1, \ldots, L_n \cup \{(B_1, A), \ldots, (B_n, A)\}$. Finally, by construction, (N, L) is unique in this sense. $\Box$

Corollary 5 Any $A \in G$ that is not a *why* argument is either in G^p or a top rule of some argument in G^p. And any rule or premise of an argument G^p is an argument in G.

Proof:
The second holds by construction of G^p, while the first follows since any A^T corresponds to an argument in G^p as specified in Proposition 4(1). $\Box$

Lemma 6 For any labelling L of a *wbd-AF* G, any $A \in G$ is labelled *in* in L iff all defeaters of A are *out* in L and for some $T \in T^A$, all children of A in T are *in* in L.

Proof:
From right to left is immediate, as is the first part from left to right. For the remainder, note that among all defeaters of A that are *out* in L are all its *why* attackers. Then each such attacker has an attacker from $\mathcal{K}$ or $\mathcal{R}_d$ that is *in*. But then by construction of T^A every set consisting of one such attacker for every *why* attacker of A is for some $T \in T^A$ the set of all children of A in T. $\Box$

Lemma 7 For any *wbd-AF* G with corresponding *pAF* G^p and all A^T and B^T in G^T it holds that B defeats A iff B^p defeats A^p for any A^p and B^p in G^p corresponding to A^T and B^T.

Proof: obvious.

Now the following correspondences can be proven between labellings of G and G^p.

Theorem 8 *Let G be a completed wbd-AF corresponding to $(AS, KB, \leq)$ with corresponding pAF G^p and let pAF' correspond to $(AS, KB, \leq)$. Then for any $\varphi \in \mathcal{L}$ it holds that φ is defensible (justified) on the basis of G iff φ is defensible (justified) on the basis of G^p iff φ is defensible (justified) on the basis of pAF'.*

Proof:
This is proven by proving the following observations:

1. Any non-why argument $A \in G$ is *in* in some labelling (all labellings) of G iff for some $T \in A^T$ the argument corresponding to T is *in* in some p-labelling (all p-labellings) of G^p.

2. Any argument $A \in G^p$ is *in* in some p-labelling (all p-labellings) of G^p iff A^T is *in* in some labelling (all labellings) of G, where A^T is the root of some $T \in T^G$ such that A corresponds to T.

3. (1) and (2) also hold if G^p is replaced by pAF'.

Then (1) and (2) follow from the structural similarity between any A^T and its corresponding argument in G^p and Proposition 4, Corollary 5 and Lemmas 6 and 7. The idea from left to right is that any labelling of G corresponds via T^G with a p-labelling of G^p, since the set of all arguments in G^p corresponding to any element of T^G is precisely the set of all arguments in G^p. For the same reason, from right to left any p-labelling of G^p corresponds via T^G to a labelling of G.
Finally, (3) follows from (1), (2) and Proposition 3. $\square$

This result is illustrated by the correspondence between the (unique) labelling of Figure 5 and the (also unique) p-labelling of Figure 4. A further illustration is given by the following example:

$p \Rightarrow_{r1} q$
$r \Rightarrow_{r2} p$
$s \Rightarrow_{r3} p$
$\Rightarrow_{r4} s$
$\Rightarrow_{r5} \neg s$
$\mathcal{K} = \{r\}$
$r_4 < r_5$

This induces the *wbd-AF* of Figure 6 on the left, with its corresponding *pAF* on the right. This *wbd-AF* has a unique labelling in all semantics, in which $p \Rightarrow_{r1} q$ is *in*, but the corresponding G^p contains two alternative arguments with top rule r_1, one of which is *in* but the other is *out* in the (unique) p-labelling of G^p. This example thus shows that in general labellings of a *wbd-AF* do not have the same meaning as (p)-labellings of *SAFs* or *pSAFS*. In a wbd-labelling, that a rule is *in* means that there is *some* acceptable way of deriving its antecedents. By contrast, in a p-labelling, that an argument is *in* means that this argument with a *specifically indicated* way of deriving the antecedents of its top rule is acceptable.

Finally, I briefly compare the approach of this section with [23]. There are some minor differences: [23] make some specific assumptions on the logical object language, which they assume to just consist of propositional literals, that is, atoms and their negations; furthermore, they replace a *why* attack on a rule antecedent by an attack with the contradictory literal. A more important difference is that [23] do not assume a knowledge base but allow atomic arguments for any literal; moreover, such literals can not only be (asymmetrically) attacked by a rule for the original literal but also (symmetrically) by the original literal. As a consequence, the focus in [23] is not on deriving arguments from a knowledge base, which makes a detailed comparison with the present approach not straightforward.

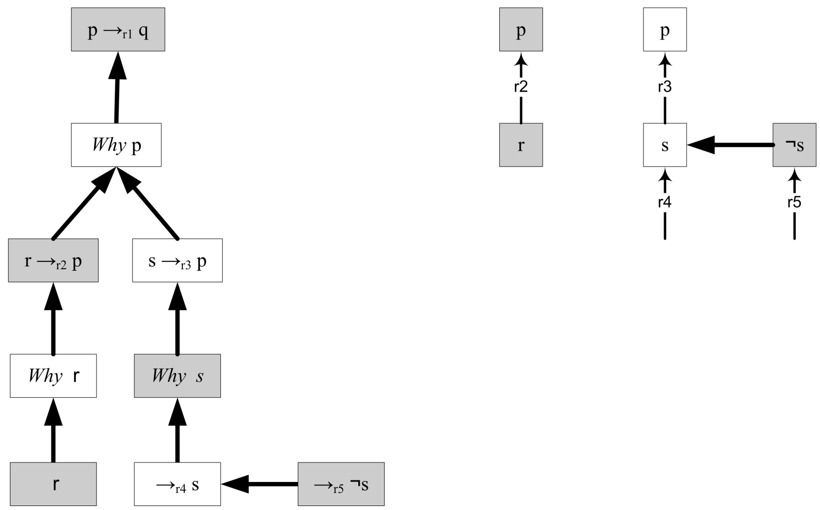

Figure 6: *wbd-AF* with corresponding *pAF*

6 Conclusion

This paper has investigated three ways to instantiate [8]'s abstract argumentation framework. The investigations have provided some reasons to believe that when the aim is to model realistic argumentation, too much focus on abstract argumentation frameworks may be harmful. First some presentational drawbacks were identified of systems for structured argumentation with inferential relations between arguments that explicitly model argument evaluation in terms of abstract argumentation frameworks. It was shown that these systems can be reformulated in a way that incorporates Dung's semantics in a stepwise evaluation of arguments, which arguably facilitates more natural explanations of dialectical argumentation evaluation. Then a method for modelling the evaluation of arguments with structure was examined in which the inferential relations between arguments are not taken into account. It turned out that this method only works well under some limiting assumptions, which do not need to be made in a model like $ASPIC^+$ that explicitly models inferential relations between arguments. Thus this paper adds to a growing body of evidence (cf. [20, 14] that adequate models of realistic argumentation should take both the structure of arguments and their inferential relations into account.

7 Afterword – some personal notes on Trevor Bench-Capon and his work

I first met Trevor Bench-Capon at JURIX 1990 in Leiden, The Netherlands, but it was not until the late 1990s that we became closer, starting with a visit by me in april that year to Liverpool, and continued during the famous Bonskeid meeting on argumentation in the Scottish hills in the summer of 2000, where we used to play an outdoors game of chess after every lunch and dinner (chess may be the only field at which I can (only slightly) top Trevor's level. Even with table tennis I suffered some terrible defeats). During my stay in Liverpool in 1999 I gave a talk in which, among other things, I summarised Dung's famous work on abstract argumentation frameworks. If I remember correctly, then neither Trevor nor his close colleague Paul Dunne knew Dung's work at the time. I still remember that especially Paul asked me many questions about Dung's work during and after my talk, and later it turned out that my visit to Liverpool had sparked a long series of AI journal publications by the two them (I am glad to see that Paul acknowledges my modest historical role in his contribution to this volume).

Over the years Trevor and I found ourselves interacting more and more during conferences and later during the ASPIC project, always to my great pleasure, because of Trevor's intelligence, his knowledge of just about everything, his wisdom and his unbeatable sense of humour. One of the great joys of discussing research with Trevor is that he often shows you another way of looking at things. However, if this way is not quite yours, then such discussions do not always spark immediate collaboration. This may be one of the reasons why we started collaborating relatively late, namely in 2006, during a new visit of me to Liverpool, now for one month.

One of the points at which we did not always fully agree was how the structure of arguments should be accounted for in the context of Dung's abstract frameworks. Although we both agreed and still agree that Dung's work is a great formal tool for the study of argumentation, I have always thought that it should be combined with accounts of the structure of arguments and their inferential relations. In recent years I have become increasingly concerned about the dominance of purely abstract research on argumentation in our field, and in some recent papers I have (partly with Sanjay Modgil) identified some pittfalls of fully abstract approaches [20, 14]. I always found it slightly unfortunate that one of Trevor's best ideas ever (the inclusion of values in models of argumentation) was first formalised at the fully abstract level. Indeed, our first joint publication [5] was an attempt to re-model Trevor's idea (as meanwhile elaborated with Katie Atkinson) in terms of an account of structured argumentation (in fact a preliminary version of the $ASPIC^+$ framework). The work on this paper was a real joy for me (I still have great memories of our exciting lunch-time conversations at Costa's), but I had the impression that Trevor was at that time not yet fully convinced that the $ASPIC^+$-like technical machinery we used is useful. In fact, during one of our lunch meetings he showed me on a napkin how the structure of arguments can be accounted for purely within Dung's abstract frameworks. At that time I found his idea quite exotic, but later I came to realise that it made perfect sense. Section 5 of this paper has been an attempt to acknowledge this by relating Trevor's idea to $ASPIC^+$.

Whatever the merits are of fully abstract approaches to argumentation, the main idea of Trevor's [4], namely that practical argumentation needs to take societal values into account, has been extremely influential in recent years, and rightly so. Moreover, Trevor now also acknowledges the benefits of combining abstract with structured fameworks for argumentation, witness, for example, our recent joint work with Adam Wyner and Katie Atkinson on formalising argument schemes for legal case-based reasoning in ASPIC+ [22]. I hope I am not too unmodest if I claim that this (besides bringing Dung's gospel to Merseyside) is one small point at which I have been able to influence Trevor's work. If so, then this is just a small and insufficient return for the many great insights that he gave me and the many enjoyable moments I could have with him. I truly hope that there will be many more of such moments to come.

References

[1] L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *International Journal of Automated Reasoning*, 29:125–169, 2002.

[2] K. Atkinson and T.J.M. Bench-Capon. Practical reasoning as presumptive argumentation using action based alternating transition systems. *Artificial Intelligence*, 171:855–874, 2007.

[3] P. Baroni and M. Giacomin. Resolution-based argumentation semantics. In Ph. Besnard, S. Doutre, and A. Hunter, editors, *Computational Models of Argument. Proceedings of COMMA 2008*, pages 25–36, Amsterdam etc, 2008. IOS Press.

[4] T.J.M. Bench-Capon. Persuasion in practical argument using value-based argumentation frameworks. *Journal of Logic and Computation*, 13:429–448, 2003.

[5] T.J.M. Bench-Capon and H. Prakken. Justifying actions by accruing arguments. In P.E. Dunne and T.B.C. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 247–258. IOS Press, Amsterdam etc, 2006.

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

[7] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171:286–310, 2007.

[8] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and n–person games. *Artificial Intelligence*, 77:321–357, 1995.

[9] T.F. Gordon, H. Prakken, and D.N. Walton. The Carneades model of argument and burden of proof. *Artificial Intelligence*, 171:875–896, 2007.

[10] N. Gorogiannis and A. Hunter. Instantiating abstract argumentation with classical-logic arguments: postulates and properties. *Artificial Intelligence*, 175:1479–1497, 2011.

[11] H. Jakobovits. *On the Theory of Argumentation Frameworks*. Doctoral dissertation Free University Brussels, 2000.

[12] H. Jakobovits and D. Vermeir. Robust semantics for argumentation frameworks. *Journal of Logic and Computation*, 9:215–261, 1999.

[13] S. Modgil. Hierarchical argumentation. In *Proceedings of the 10th European Conference on Logics in Artificial Intelligence (JELIA)*, pages 319–332, 2006.

[14] S. Modgil and H. Prakken. Resolutions in structured argumentation. In B. Verheij, S. Woltran, and S. Szeider, editors, *Computational Models of Argument. Proceedings of COMMA 2012*, pages 310–321. IOS Press, Amsterdam etc, 2012.

[15] S. Modgil and H. Prakken. A general account of argumentation with preferences. *Artificial Intelligence*, 195:361–397, 2013.

[16] D. Nute. Defeasible logic. In D. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, pages 253–395. Clarendon Press, Oxford, 1994.

[17] J.L. Pollock. *Cognitive Carpentry. A Blueprint for How to Build a Person*. MIT Press, Cambridge, MA, 1995.

[18] H. Prakken. Coherence and flexibility in dialogue games for argumentation. *Journal of Logic and Computation*, 15:1009–1040, 2005.

[19] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1:93–124, 2010.

[20] H. Prakken. Some reflections on two current trends in formal argumentation. In *Logic Programs, Norms and Action. Essays in Honour of Marek J. Sergot on the Occasion of his 60th Birthday*, pages 249–272. Springer, Berlin/Heidelberg, 2012.

[21] H. Prakken and G. Sartor. On modelling burdens and standards of proof in structured argumentation. In K.D. Atkinson, editor, *Legal Knowledge and Information Systems. JURIX 2011: The Twenty-fourth Annual Conference*, pages 83–92. IOS Press, Amsterdam etc., 2011.

[22] H. Prakken, A.Z. Wyner, T.J.M. Bench-Capon, and K.D. Atkinson. A formalisation of argumentation schemes for legal case-based reasoning in ASPIC+. *Journal of Logic and Computation*, 2013. In press.

[23] A.Z. Wyner, T.J.M. Bench-Capon, and P.E. Dunne. Instantiating knowledge bases in abstract argumentation frameworks. In *Proceedings of the AAAI Fall 2009 Symposium on The Uses of Computational Argumentation*, pages 76–83. AAAI Press, 2009. Technical Report SS-09-06.