

## RESEARCH ARTICLE

### The *ASPIC*<sup>+</sup> framework for structured argumentation: a tutorial

Sanjay Modgil<sup>a</sup> and Henry Prakken<sup>b</sup>

<sup>a</sup>*Department of Informatics, King's College London, UK;* <sup>b</sup>*Department of Information and Computing Sciences, Utrecht University & Faculty of Law, University of Groningen, The Netherlands*

(Received 00 Month 200x; final version received 00 Month 200x)

This article gives a tutorial introduction to the *ASPIC*<sup>+</sup> framework for structured argumentation. The philosophical and conceptual underpinnings of *ASPIC*<sup>+</sup> are discussed, the main definitions are illustrated with examples and several ways are discussed to instantiate the framework and to reconstruct other approaches as special cases of the framework. The *ASPIC*<sup>+</sup> framework is based on two ideas: the first is that conflicts between arguments are often resolved with explicit preferences, and the second is that arguments are built with two kinds of inference rules: strict, or deductive rules, whose premises guarantee their conclusion, and defeasible rules, whose premises only create a presumption in favour of their conclusion. Accordingly, arguments can in *ASPIC*<sup>+</sup> be attacked in three ways: on their uncertain premises, or on their defeasible inferences, or on the conclusions of their defeasible inferences. *ASPIC*<sup>+</sup> is not a system but a framework for specifying systems. A main objective of the study of the *ASPIC*<sup>+</sup> framework is to identify conditions under which instantiations of the framework satisfy logical consistency and closure properties.

#### 1. Introduction

This paper presents a tutorial introduction to the *ASPIC*<sup>+</sup> framework for structured argumentation. *ASPIC*<sup>+</sup> is meant to generate abstract argumentation frameworks in the sense of Dung (1995). Such frameworks are simply directed graphs in which the arguments (nodes) are related to other arguments by attack or defeat relations (arcs). A ‘calculus of opposition’ is then applied to a framework to determine sets of acceptable arguments (extensions). While this abstract calculus is an indispensable component of theories of argumentation, it says nothing about the structure of arguments or the nature of attack or defeat, and so provides no guidance for the modelling of actual argumentation problems. That is to say, how should one model an argument’s constituent reasons and the inferential steps from these reasons to an argument’s claim, how do these structural properties of arguments determine attacks between arguments, and in turn, how should preferences be used to determine whether one argument successfully attacks (defeats) another argument? Moreover, at the fully abstract level one cannot study properties that one would intuitively expect to hold of the arguments in extensions; for example that their claims should be mutually consistent or that if the claims of arguments in an extension deductively imply some formula, then there should be an argument in that extension that claims that formula (i.e., the arguments are closed under deductive inference).

The question then arises as to whether one can give a general structured account of argumentation that is intermediate in its level of abstraction between concrete logics and the fully abstract level, providing guidance on the structure of arguments, the nature of attacks, and the use of preferences, while at the same time accommodating a broad range of instantiating logics and allowing for the study of conditions under which the various desirable properties are satisfied by these instantiations. The *ASPIC*<sup>+</sup> framework is an attempt to answer this question in the affirmative.

Historically, the *ASPIC*<sup>+</sup> framework originates from the European ASPIC project that

ran from 2004 – 2007, which (among other things) aimed to integrate and consolidate the main approaches to structured argumentation. The first relevant publication was Caminada and Amgoud (2007), which also introduced the idea of desirable properties (rationality postulates) for structured argumentation. The *ASPIC*<sup>+</sup> framework as published in this paper was generalised and extended in Prakken (2010), Modgil and Prakken (2013) so as to capture a broader range of instantiating concrete logics (e.g., Besnard and Hunter (2009)), other structured general accounts of argumentation (Amgoud and Besnard (2009), Bondarenko et al. (1997)), and to account for the use of preferences to decide which attacks succeed as defeats. Unless indicated otherwise, the version of *ASPIC*<sup>+</sup> presented here is the one of Modgil and Prakken (2013).

The remainder of this tutorial paper is organised as follows. In Section 2 we motivate the design choices for *ASPIC*<sup>+</sup>; that is to say *ASPIC*<sup>+</sup>'s structural account of arguments, the nature of attacks implied by this account, and the use of preferences. Section 3 then presents a formal account of *ASPIC*<sup>+</sup> together with illustrative examples. This section also elaborates on the two key ideas that underpin *ASPIC*<sup>+</sup>, namely that conflicts between arguments are sometimes resolved with explicit preferences, and that arguments are built with two kinds of inference rules: strict rules, which logically entail their conclusion, and defeasible rules, which only create a presumption in favour of their conclusion. Sections 4 and 5 then review specific uses of *ASPIC*<sup>+</sup>, both as a means to define new argumentation logics and as a framework for translating and studying existing approaches.

## 2. *ASPIC*<sup>+</sup> Design choices and Overview

People argue to remove doubt about a claim (Walton 2006, p. 1), by giving reasons why one should accept the claim and by defending these reasons against criticism. The strongest way to remove doubt is to show that the claim deductively follows from indisputable grounds. A mathematical proof from the axioms of arithmetic is like this: its grounds are mathematical axioms, while its inferences are deductively sound. So such a proof cannot be attacked in any way: not on its grounds and not on its inferences. However, such perfection is not attainable in real life: our grounds may not be indisputable or they may provide less than conclusive support for their claim.

Suppose we believe that John was in Holland Park some morning and that Holland Park is in London. Then we can deductively reason from these beliefs, to conclude that John was in London that morning. So the reasoning cannot be attacked. However, perfection remains unattainable since the argument is still fallible: its grounds may turn out to be wrong. For instance, Jan may tell us that he met John in Amsterdam that morning around the same time. We now have a reason against our belief that John was in Holland Park that morning, since witnesses usually speak the truth. Can we retain our belief or must we give it up? The answer to this question determines whether we can accept that John was in London that morning.

Maybe we originally believed that John was in Holland Park for a reason. Maybe we went jogging in Holland Park and we saw John. We then have a reason supporting our belief that John was in Holland Park that morning, since we know that our senses are usually accurate. But we cannot be sure, since Jan told us that he met John in Amsterdam that morning around the same time. Perhaps our senses betrayed us this morning? But then we hear that Jan has a reason to lie, since John is a suspect in a robbery in Holland Park that morning and Jan and John are friends. We then conclude that the basis for questioning our belief that John was in Holland Park that morning (namely, that witnesses usually speak the truth and Jan witnesses John in Amsterdam) does not apply to witnesses who have a reason to lie. So our reason in support of our belief is undefeated and we accept it.

If we want to formalise a logic for argumentation, then this simple example already suggests a number of issues we have to deal with. At least two further important design

decisions have to be made: how can arguments be built, i.e., how can claims be supported with grounds, and how can arguments be attacked? We shall see that the answers to these two questions are related.

First, the claims and beliefs in our example were supported in various ways: in the first case we appealed to the principles of deductive inference when concluding that John was in London. *ASPIC*<sup>+</sup> is therefore designed so that arguments can be constructed using deductive or *strict* inference rules that license deductive inferences from premises to conclusions. However, in the other two cases the reasoning from grounds to claim appealed to the reliability of, respectively, our senses and witnesses as sources of information. Should these kinds of support (inferences) from grounds to claims be modelled as deductive?

To help answer this question, consider that our informal example contains three ways of attacking an argument: 1) Our initial argument that John was in London was attacked by the witness argument on its ground, or *premise*, that John was in Holland Park that morning; 2) We then modified our initial argument by extending it with an additional argument for the attacked premise, but the extended argument was still attacked (by the witness argument) on the (now) intermediate conclusion that John was in Holland Park that morning; 3) Finally, we counterattacked the witness argument not on a premise or conclusion but on the reasoning from the grounds to the claim: namely, the inference step from the premise that Jan said he met John in Amsterdam that morning to the claim that John was in Amsterdam that morning (note that here we regard the principle that witnesses usually speak the truth as an inference rule).

Now, returning to the question whether all kinds of inference should be deductive, the second type of attack would not be possible on the deductively inferred intermediate conclusion since the nature of deductive support is that it is absolutely watertight: if one accepts all antecedents of a deductively valid inference rule, then one must also accept its consequent no matter what, on the penalty of being irrational. If the antecedents of a deductively valid inference rule are true, then its consequent must also be true. So if we have reason to believe that the conclusion of a deductive inference is not true, then there must be something wrong with its premises (which may in turn be the conclusions of sub-arguments). It is for this very same reason that the third type of attack, on the deductive inferential step itself, is also not possible.

*ASPIC*<sup>+</sup> is therefore designed to comply with the common-sense and philosophically argued position (Pollock (1995, p.41); Pollock (2009, p. 173)) advocating the rationality of supporting claims with grounds that do not deductively entail them. In other words, the fallibility of an argument need not only be located in its premises, but can also be located in the inference steps from premises to conclusion. Thus, arguments in *ASPIC*<sup>+</sup> can be constructed using *defeasible* inference rules, and arguments can be attacked on the application of such defeasible inference rules, in keeping with the interpretation that the premises of such a rule presumptively, rather than deductively, support their conclusions,

However, some would argue that the second and third type of attacks can be simulated using only deductive rules (specifically the deductive rules of classical logic) by augmenting the antecedents of these rules with normality premises. For example, with regard to the second type of attack, could we not say that our argument claiming that John was in Holland Park that morning since we saw him there has an implicit premise *our senses functioned normally*, and that the argument that John was in Amsterdam that morning in fact attacks this implicit premise, rather than its claim, thus reducing attacks on conclusions to attacks on premises? With regard to the third type of attack, could we not say that instead of attacking the defeasible inference step from Jan's testimony to the claim that John was in Amsterdam, we could model this step as deductive, and then add the premise that normally witnesses speak the truth, and then direct the attack at this premise? In other words, can we reduce attacks on inferences to attacks on premises?

In answer to these questions, we first note that some have argued that such deductive

simulations are prone to yielding counterintuitive results. This is a topic that we will return to and examine in more detail in Section 4.5. Second, we claim that there is some merit in modelling the everyday practice of ‘jumping to defeasible conclusions’ and of considering arguments for contradictory conclusions. This is especially important given that one of the argumentation paradigm’s key strengths is its characterisation of formal logical modes of reasoning in a way that corresponds with human modes of reasoning and debate.

The above discussion introduced the notion of *fallible* premises that can be attacked. However  $ASPIC^+$  also wants to allow you to distinguish premises that are axiomatic and so cannot be attacked. We discuss the uses of such premises in Section 4, but for the moment we can summarise by saying that  $ASPIC^+$  arguments can be constructed from fallible and infallible premises (respectively called *ordinary* and *axiom* premises in Section 3), and strict and defeasible inference rules, and that arguments can be attacked on their ordinary premises, the conclusions of defeasible inference rules, and the defeasible inference steps themselves. Finally, a key feature of the  $ASPIC^+$  framework is that it accommodates the use of preferences over arguments, so that an attack from one argument to another only succeeds (as a defeat) if the attacked argument is not stronger than (strictly preferred to) the attacking argument, according to some given preference relation. The justified  $ASPIC^+$  arguments are then evaluated with respect to the Dung framework relating  $ASPIC^+$  arguments by the defeat relation. Since requirements for use of preferences in argumentation (and more generally for conflict resolution in non-monotonic logics) are well established in the literature, we will not here justify the need for preferences. However, examples are given in the remainder of the paper.

### 3. The framework defined: Special case with ‘ordinary’ negation

In what follows we briefly review Dung’s abstract argumentation theory, and then the  $ASPIC^+$  framework. Note that in this section we present a special case of  $ASPIC^+$ , in which conflict is based on the standard classical notion of negation, and then in Section 5 we replace negation by a more general notion of conflict between formulae.

#### 3.1. Abstract argumentation

As briefly discussed in Section 1,  $ASPIC^+$  was explicitly designed to be intermediate in abstraction between concrete logics and Dung’s abstract argumentation frameworks; the idea being that logics conforming to the  $ASPIC^+$  framework define  $ASPIC^+$  arguments and defeats that then comprise a Dung framework. We therefore first briefly review the main concepts from Dung’s abstract argumentation theory.

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*. We say that  $A$  *strictly defeats*  $B$  if  $A$  defeats  $B$  while  $B$  does not defeat  $A$ . A semantics for  $AF$ s returns sets of arguments called *extensions*, which are internally coherent and defend themselves against attack.

**Definition 3.1:** Let  $(\mathcal{A}, \mathcal{D})$  be a  $AF$ . For any  $X \in \mathcal{A}$ ,  $X$  is acceptable with respect to some  $S \subseteq \mathcal{A}$  iff  $\forall Y$  s.t.  $(Y, X) \in \mathcal{D}$  implies  $\exists Z \in S$  s.t.  $(Z, Y) \in \mathcal{D}$ . Let  $S \subseteq \mathcal{A}$  be *conflict free*, i.e., there are no  $A, B$  in  $S$  such that  $(A, B) \in \mathcal{D}$ . Then:

- $S$  is an *admissible* extension iff  $X \in S$  implies  $X$  is acceptable w.r.t.  $S$ ;
- $S$  is a *complete* extension iff  $X \in S$  whenever  $X$  is acceptable w.r.t.  $S$ ;
- $S$  is a preferred extension iff it is a set inclusion maximal complete extension;
- $S$  is the grounded extension iff it is the set inclusion minimal complete extension;
- $S$  is a stable extension iff it is preferred and  $\forall Y \notin S, \exists X \in S$  s.t.  $(X, Y) \in \mathcal{D}$ .

For  $T \in \{\text{complete, preferred, grounded, stable}\}$ ,  $X$  is *sceptically* or *credulously* justified under the  $T$  semantics if  $X$  belongs to all, respectively at least one,  $T$  extension.

### 3.2. Argumentation systems, knowledge bases, and arguments

To use  $ASPIC^+$ , you need to provide the following information. You must choose a *logical language*  $\mathcal{L}$  closed under negation  $\neg$  (which we later replace with a more general notion of conflict). You must then provide two (possibly empty) sets of *strict* ( $\mathcal{R}_s$ ) and *defeasible* ( $\mathcal{R}_d$ ) inference rules. If you provide a non-empty set of defeasible rules, you then need to also specify which well-formed formulas in  $\mathcal{L}$  correspond to (i.e., name) which defeasible rule in  $\mathcal{R}_d$ . To do the latter requires specifying a partial function  $n$  from  $\mathcal{R}_d$  to  $\mathcal{L}$ . These names can then be used when attacking arguments on defeasible inference steps. Informally,  $n(r)$  is a wff in  $\mathcal{L}$  which says that the defeasible rule  $r \in \mathcal{R}$  is applicable, so that an argument claiming  $\neg n(r)$  attacks the inference step in the corresponding rule<sup>1</sup>.

The above is summarised in the following formal definition:

**Definition 3.2:** [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  $\varphi_1, \dots, \varphi_n \rightarrow \varphi$  and  $\varphi_1, \dots, \varphi_n \Rightarrow \varphi$  respectively (where  $\varphi_i, \varphi$  are meta-variables ranging over wff in  $\mathcal{L}$ ), and  $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$ .
- $n$  is a partial function such that  $n : \mathcal{R}_d \rightarrow \mathcal{L}$ .

We write  $\psi = -\varphi$  just in case  $\psi = \neg\varphi$  or  $\varphi = \neg\psi$  (we will sometimes informally say that formulas  $\varphi$  and  $-\varphi$  are each other's negation).

It is important to stress here that  $ASPIC^+$ 's strict and defeasible inference rules are *not* object-level formulae in the language  $\mathcal{L}$ , but are meta to the language, allowing one to deductively, respectively defeasibly, infer the rule's consequent from the rule's antecedents. Such inference rules may range over arbitrary formulae in the language, in which case they will, as usual in logic, be specified as *schemes*. For example, a scheme for strict inference rules capturing modus ponens for the material implication of classical logic can be written as  $\alpha, \alpha \supset \beta \rightarrow \beta$ , where  $\alpha$  and  $\beta$  are metavariables for wff in  $\mathcal{L}$ . Alternatively, strict or defeasible inference rules may be domain-specific in that they reference specific formulae, as in the defeasible inference rule concluding that an individual flies if that individual is a bird:  $Bird \Rightarrow Flies$ . We will further discuss these distinct uses of inference rules in Section 4.

If you want to use  $ASPIC^+$ , then an argumentation system is not all you have to specify: you must also specify from which body of information the premises of an argument can be taken. We call this a knowledge base, and as discussed in Section 2, distinguish ordinary premises, which are uncertain and so can be attacked, and premises that are axioms, hence certain, and so cannot be attacked.

**Definition 3.3:** [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*).

$ASPIC^+$  leaves you fully free to choose your language, what is an axiom and what is an ordinary premise and how you specify your strict and defeasible rules. However some care needs to be taken in making these choices, to ensure that the result of argumentation is guaranteed to be well-behaved. By 'well-behaved' we mean that the desirable properties

<sup>1</sup> $n$  is a partial function since you may want to enforce that some defeasible inference steps cannot be attacked.

<sup>2</sup>In this paper we use  $\supset$  to denote the material implication connective of classical logic.

proposed by Caminada and Amgoud (2007) are satisfied; for example, that the conclusions of arguments in the same extension are mutually consistent (we will define below what this means) and are closed under application of strict inference rules (whatever you can derive from your conclusions of arguments in an extension, with strict rules alone, is already a conclusion of an argument in that extension). In Section 4 we present some theorems which tell you how you can make your choices in such a way that the result is guaranteed to be well-behaved. These theorems will talk about two notions of consistency, namely, direct and indirect consistency. Indirect consistency is defined in terms of the closure of a set of well-formed formulas under application of strict inference rules. Informally, the strict closure of a set of wff is the set itself plus everything that can be derived from it when only applying strict rules.

**Definition 3.4:** For any  $S \subseteq \mathcal{L}$ , let the closure of  $S$  under strict rules, denoted  $Cl(S)$ , be the smallest set containing  $S$  and the consequent of any strict rule in  $\mathcal{R}_s$  whose antecedents are in  $Cl(S)$ . Then a set  $S \subseteq \mathcal{L}$  is

- *directly consistent* iff  $\nexists \psi, \varphi \in S$  such that  $\psi = -\varphi$
- *indirectly consistent* iff  $Cl(S)$  is directly consistent.

We call the combination of an argumentation system and a knowledge base an argumentation theory:

**Definition 3.5:** [Argumentation theory] An *argumentation theory* is a tuple  $AT = (AS, \mathcal{K})$  where  $AS$  is an argumentation system and  $\mathcal{K}$  is a knowledge base in  $AS$ .

$ASPIC^+$  arguments are now defined relative to an argumentation theory  $AT = (AS, \mathcal{K})$ , and chain applications of the inference rules from  $AS$  into inference trees, starting with elements from the knowledge base  $\mathcal{K}$ . In what follows, for a given argument, the function Prem returns all the formulas of  $\mathcal{K}$  (called *premises*) used to build the argument, Conc returns its conclusion, Sub returns all its sub-arguments, DefRules returns all the defeasible rules of the argument and TopRule returns the last inference rule used in the argument.

**Definition 3.6:** [Argument] An *argument*  $A$  on the basis of an argumentation theory with a knowledge base  $\mathcal{K}$  and an argumentation system  $(\mathcal{L}, \mathcal{R}, n)$  is:

- (1)  $\varphi$  if  $\varphi \in \mathcal{K}$  with:  $Prem(A) = \{\varphi\}$ ,  $Conc(A) = \varphi$ ,  $Sub(A) = \{\varphi\}$ ,  $DefRules(A) = \emptyset$ ,  $TopRule(A) = \text{undefined}$ .
- (2)  $A_1, \dots, A_n \rightarrow \psi$  if  $A_1, \dots, A_n$  are arguments such that there exists a strict rule  $Conc(A_1), \dots, Conc(A_n) \rightarrow \psi$  in  $\mathcal{R}_s$ .  
 $Prem(A) = Prem(A_1) \cup \dots \cup Prem(A_n)$ ,  
 $Conc(A) = \psi$ ,  
 $Sub(A) = Sub(A_1) \cup \dots \cup Sub(A_n) \cup \{A\}$ .  
 $DefRules(A) = DefRules(A_1) \cup \dots \cup DefRules(A_n)$ ,  
 $TopRule(A) = Conc(A_1), \dots, Conc(A_n) \rightarrow \psi$
- (3)  $A_1, \dots, A_n \Rightarrow \psi$  if  $A_1, \dots, A_n$  are arguments such that there exists a defeasible rule  $Conc(A_1), \dots, Conc(A_n) \Rightarrow \psi$  in  $\mathcal{R}_d$ .  
 $Prem(A) = Prem(A_1) \cup \dots \cup Prem(A_n)$ ,  
 $Conc(A) = \psi$ ,  
 $Sub(A) = Sub(A_1) \cup \dots \cup Sub(A_n) \cup \{A\}$ ,  
 $DefRules(A) = DefRules(A_1) \cup \dots \cup DefRules(A_n) \cup \{Conc(A_1), \dots, Conc(A_n) \Rightarrow \psi\}$ ,  
 $TopRule(A) = Conc(A_1), \dots, Conc(A_n) \Rightarrow \psi$ .

**Example 3.7** Consider a knowledge base in an argumentation system with  $\mathcal{L}$  consisting of  $p, q, r, s, t, u, v, w, x, d_1, d_2, d_3, d_4, d_5, d_6$  and their negations, with  $\mathcal{R}_s = \{s_1, s_2\}$  and  $\mathcal{R}_d = \{d_1, d_2, d_3, d_4, d_5, d_6\}$ , where

$$\begin{array}{lll}
d_1: p \Rightarrow q & d_4: u \Rightarrow v & s_1: p, q \rightarrow r \\
d_2: s \Rightarrow t & d_5: v, x \Rightarrow \neg t & s_2: v \rightarrow \neg s \\
d_3: t \Rightarrow \neg d_1 & d_6: s \Rightarrow \neg p &
\end{array}$$

Moreover,  $\mathcal{K}_n = \{p\}$  and  $\mathcal{K}_p = \{s, u, x\}$ . Note that in presenting the example, we have informally used names  $d_i$  to refer to defeasible inference rules. We now define the  $n$  function that formally assigns wff  $d_i$  to such rules, i.e., for any rule informally referred to as  $d_i$ , we have that  $n(d_i) = d_i$ , so that ‘ $n(d_1) = d_1$ ’ is a shorthand for  $n(p \Rightarrow q) = d_1$ . In further examples we will often specify the  $n$  function in the same way.<sup>1</sup>

An argument for  $r$  (i.e., with conclusion  $r$ ) is displayed in Figure 1, with the premises at the bottom and the conclusion at the top of the tree. In this and the next figure, the type of a premise is indicated with a superscript and defeasible inferences, underminable premises and rebuttable conclusions are displayed with dotted lines. The figure also displays the formal structure of the argument. We have that

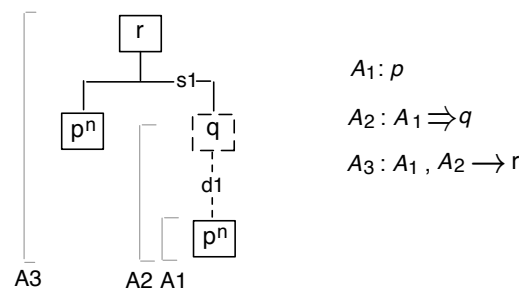


Figure 1. An argument

$$\begin{array}{ll}
\text{Prem}(A_3) = \{p\} & \text{DefRules}(A_3) = \{d_1\} \\
\text{Conc}(A_3) = r & \text{TopRule}(A_3) = s_1 \\
\text{Sub}(A_3) = \{A_1, A_2, A_3\} &
\end{array}$$

The distinction between two kinds of inference rules and two kinds of premises motivates a distinction into four kinds of arguments.

**Definition 3.8:** [Argument properties] An argument  $A$  is *strict* if  $\text{DefRules}(A) = \emptyset$ ; *defeasible* if  $\text{DefRules}(A) \neq \emptyset$ ; *firm* if  $\text{Prem}(A) \subseteq \mathcal{K}_n$ ; *plausible* if  $\text{Prem}(A) \cap \mathcal{K}_p \neq \emptyset$ . We write  $S \vdash \varphi$  if there exists a strict argument for  $\varphi$  with all premises taken from  $S$ , and  $S \sim \varphi$  if there exists a defeasible argument for  $\varphi$  with all premises taken from  $S$ .

**Example 3.9** In Example 3.7 the argument  $A_1$  is both strict and firm, while  $A_2$  and  $A_3$  are defeasible and firm. Furthermore, we have that  $\mathcal{K} \vdash p$ ,  $\mathcal{K} \sim q$  and  $\mathcal{K} \sim r$ .

### 3.3. Attack and defeat

Recall that  $\text{ASPIC}^+$  is meant to generate Dung-style abstract argumentation frameworks, that is, a set of arguments with a binary relation of defeat. Having defined arguments above, we now define the attack relation and then, as discussed in Section 2, we apply preferences to determine the defeat relation (in fact Dung called his relation ‘‘attack’’ but we reserve this term for the basic notion of conflict, to which we then apply preferences).

<sup>1</sup>In our further examples we will often leave the logical language  $\mathcal{L}$  and the  $n$  function implicit, trusting that they will be obvious.

### 3.3.1. Attack

We now first present the three ways in which arguments in  $ASPIC^+$  can be in conflict, that is, three kinds of attack. In short, arguments can be attacked on a conclusion of a defeasible inference (rebutting attack), on a defeasible inference step itself (undercutting attack), or on an ordinary premise (undermining attack). As discussed in Section 2, that arguments cannot be attacked on their strict inferences goes without saying. We also discussed why arguments cannot be attacked on the conclusions of strict inferences: if the antecedents of a deductively valid inference rule are true, then its consequent must also be true no matter what. So if we have reason to believe that the conclusion of a deductive inference is not true, then there must be something wrong with the claims from which it is drawn. In Section 4.2 we will give a second reason why arguments cannot be attacked on conclusions of strict inferences. In short, this is because if we allow such attacks, then consistency and strict closure of conclusions cannot be guaranteed.

To define undercutting attack, the function  $n$  of an  $AS$  is used, which assigns to elements of  $\mathcal{R}_d$  a well-formed formula in  $\mathcal{L}$ . Recall that informally,  $n(r)$  (where  $r \in R_d$ ) means that  $r$  is applicable. Then an argument using  $r$  is undercut by any argument with conclusion  $\neg n(r)$ .

**Definition 3.10:** [attacks]  $A$  attacks  $B$  iff  $A$  undercuts, rebuts or undermines  $B$ , where:

- $A$  undercuts argument  $B$  (on  $B'$ ) iff  $\text{Conc}(A) = \neg n(r)$  for some  $B' \in \text{Sub}(B)$  such that  $B'$ 's top rule  $r$  is defeasible.
- $A$  rebuts argument  $B$  (on  $B'$ ) iff  $\text{Conc}(A) = \neg\varphi$  for some  $B' \in \text{Sub}(B)$  of the form  $B''_1, \dots, B''_n \Rightarrow \varphi$ .
- Argument  $A$  undermines  $B$  (on  $\varphi$ ) iff  $\text{Conc}(A) = \neg\varphi$  for an ordinary premise  $\varphi$  of  $B$ .

This definition allows for a distinction between direct and indirect attack: an argument can be indirectly attacked by directly attacking one of its proper subarguments. This distinction will turn out to be crucial for a proper application of preferences to resolve attacks.

**Example 3.11** In our running example argument  $A_3$  cannot be undermined, since all its premises are axioms.  $A_3$  can potentially be rebutted on  $A_2$ , with an argument for  $\neg q$ . However, the argumentation theory of our example does not allow the construction of such a rebuttal. Likewise,  $A_3$  can potentially be undercut on  $A_2$ , with an argument for  $\neg d_1$ . Our example does allow the construction of such an undercutter, namely:

$$\begin{aligned} B_1: & s \\ B_2: & B_1 \Rightarrow t \\ B_3: & B_2 \Rightarrow \neg d_1 \end{aligned}$$

Argument  $B_3$  has an ordinary premise  $s$ , so it can be undermined on  $B_1$  with an argument for  $\neg s$ :

$$\begin{aligned} C_1: & u \\ C_2: & C_1 \Rightarrow v \\ C_3: & C_2 \rightarrow \neg s \end{aligned}$$

Note that since  $C_3$  has a strict top rule, argument  $B_1$  does not in turn rebut  $C_3$ .

Argument  $B_3$  can potentially be rebut or undercut on either  $B_2$  or  $B_3$ , since both of these subarguments of  $B_3$  have a defeasible top rule. Our argumentation theory only allows for a rebutting attack on  $B_2$ :

$$\begin{aligned} C_1: & u \\ C_2: & C_1 \Rightarrow v \\ D_3: & x \\ D_4: & C_2, D_3 \rightarrow \neg t \end{aligned}$$



All relevant arguments and attacks are displayed in Figure 2.

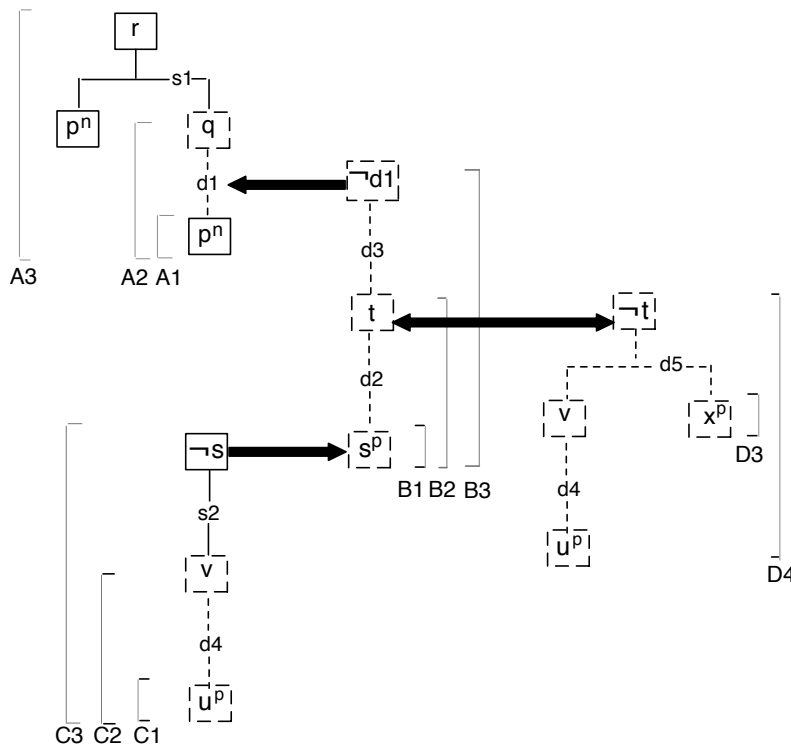


Figure 2. An argument

### 3.3.2. Defeat

The attack relation tells us which arguments are in conflict with each other: if two arguments are in conflict then they cannot both be justified. However, Definition 3.1’s notion of the acceptability of arguments is based on the notion that one argument can be used as a counter-argument to another. In general, an argument  $A$  can be used as a counter-argument to  $B$ , if  $A$  *successfully attacks*, i.e., defeats,  $B$ . Whether an attack from  $A$  to  $B$  (on its sub-argument  $B'$ ) succeeds as a defeat, may depend on the relative strength of  $A$  and  $B'$ , i.e., whether  $B'$  is *strictly stronger than*, or *strictly preferred to*  $A$ . Note that only the success of undermining and rebutting attacks is contingent on preferences; undercutting attacks succeed as defeats independently of any preferences (see Modgil and Prakken (2013) for a discussion as to why this is the case).

Where do these preferences come from? Again,  $ASPIC^+$  allows you to make any choice you like. All that  $ASPIC^+$  as a framework wants is that you as a user give a binary ordering  $\preceq$  on the set of all arguments that can be constructed on the basis of an argumentation theory. Then, as usual, if  $A \preceq B$  and  $B \not\preceq A$  then  $B$  is strictly preferred to  $A$  (denoted  $A \prec B$ ). Also, if  $A \preceq B$  and  $B \preceq A$  then  $A \approx B$ . We will later identify some conditions under which argument orderings are well-behaved in that they promote consistency and strict closure of conclusions. We will also define two example argument orderings that satisfy these conditions. However, for now all we need for defining  $ASPIC^+$ ’s defeat relation is the attack relation and a preference ordering over arguments.

How should the preference ordering be applied to resolve attacks? At first sight, it would seem that  $ASPIC^+$  can be taken to generate a so-called *preference-based argumentation framework* (PAF) in the sense of Amgoud and Cayrol (2002), that is, a triple consisting of the set of arguments, the attack relation and the argument ordering. That  $A$  defeats

$B$  could then be defined as  $A$  attacks  $B$  and  $A \not\prec B$ . However, this does not work, for two reasons. First, PAFs do not recognise that undercutting attacks succeed irrespective of preferences. More seriously, PAFs cannot express how and at which points arguments attack each other, and yet this is crucial for a proper application of preferences to attack relations. Modgil and Prakken (2013), Prakken (2012) have shown that the use of PAFs leads to violation of the rationality postulates of subargument closure and consistency (see further Section 4.2 below) in cases where  $ASPIC^+$  with the following definition satisfies these postulates.

**Definition 3.12:** [Successful rebuttal, undermining and defeat]

- $A$  *successfully rebuts*  $B$  if  $A$  rebuts  $B$  on  $B'$  and  $A \not\prec B'$ .
- $A$  *successfully undermines*  $B$  if  $A$  undermines  $B$  on  $\varphi$  and  $A \not\prec \varphi$ .
- $A$  *defeats*  $B$  iff  $A$  undercuts or successfully rebuts or successfully undermines  $B$ .

The success of rebutting and undermining attacks thus involves comparing the conflicting arguments at the points where they conflict; that is, by comparing those arguments that are in a *direct* rebutting or undermining relation with each other. The definition of successful undermining exploits the fact that an argument premise is also a subargument.

**Example 3.13** In our running example two argument orderings are relevant for whether attacks are successful: between  $B_1$  and  $C_3$  and and between  $B_2$  and  $D_4$ . Note that the undercutting attack of  $B_3$  on  $A_2$  (and thereby on  $A_3$ ) succeeds as a defeat irrespective of the argument ordering between  $B_3$  and  $A_2$ . The undermining attack of  $C_3$  on  $B_1$  succeeds if  $C_3 \not\prec B_1$ . If  $B_2 \approx D_4$  or their relation is undefined then these two arguments defeat each other, while  $D_4$  strictly defeats  $B_3$ . If  $D_4 \prec B_2$  then  $B_2$  strictly defeats  $D_4$  while if  $B_2 \prec D_4$  then  $D_4$  strictly defeats both  $B_2$  and  $B_3$ .

Let us now put all these elements together; that is the arguments and attacks defined on the basis of an argumentation theory, and a preference ordering over the arguments:

**Definition 3.14:** 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 smallest set of all finite arguments constructed from  $KB$  in  $AS$  satisfying Definition 3.6;
- $\preceq$  is an ordering on  $\mathcal{A}$ ;
- $(X, Y) \in \mathcal{C}$  iff  $X$  attacks  $Y$ .

**Example 3.15** In our running example  $\mathcal{A} = \{A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, C_3, D_3, D_4\}$ , while  $\mathcal{C}$  is such that  $B_3$  attacks both  $A_2$  and  $A_3$ , argument  $C_3$  attacks all of  $B_1, B_2, B_3$ , argument  $D_4$  attacks both  $B_2$  and  $B_3$  and, finally,  $B_2$  attacks  $D_4$ .

### 3.4. Generating Dung-style abstract argumentation frameworks

We are now ready to instantiate a Dung framework with  $ASPIC^+$  arguments and the  $ASPIC^+$  defeat relation.

**Definition 3.16 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  $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ .

The justified arguments of the above defined AF are then defined under various semantics, as in Definition 3.1<sup>1</sup>. Now one way to define the justification status of a statement is

<sup>1</sup>Note that in Modgil and Prakken (2013) we motivate the use of the  $ASPIC^+$  attack relation to define conflict-free sets, and then only use the  $ASPIC^+$  defeat relation to determine the acceptability of arguments. It turns out that under certain conditions, this way of evaluating the status of arguments is equivalent to Definition 3.1's use of the defeat relation for *both*

as follows:

**Definition 3.17:** A wff  $\varphi \in \mathcal{L}$  is *sceptically justified* if  $\varphi$  is the conclusion of a sceptically justified argument, and *credulously justified* if  $\varphi$  is not sceptically justified and is the conclusion of a credulously justified argument.

**Example 3.18** In our running example, if  $D_4$  strictly defeats  $B_2$ , then we have a unique extension in all semantics which at least contains the set  $S = \{A_1, A_2, A_3, C_1, C_2, C_3, D_3, D_4\}$ . If in addition  $C_3$  does not defeat  $B_1$ , then the extension also contains  $B_1$ . In both cases this yields that wff  $r$  is sceptically justified.

Alternatively, if  $B_2$  strictly defeats  $D_4$ , then the status of  $r$  depends on whether  $C_3$  defeats  $B_1$ . If it does, then we again have a unique extension in all semantics consisting of the set  $S$ , so  $r$  is sceptically justified. By contrast, if  $C_3$  does not defeat  $B_1$ , we obtain a unique extension with  $A_1, B_1, B_2, B_3, C_1, C_2, C_3$  and  $D_3$ , so  $r$  is neither sceptically nor credulously justified.

Finally, if  $B_2$  and  $D_4$  defeat each other, then the outcome again depends on whether  $C_3$  defeats  $B_1$ . If it does, then the situation is as in the previous case – a unique extension  $S$  – but if  $C_3$  does not defeat  $B_1$ , then the grounded extension consists of  $A_1, B_1, C_1$ - $C_3, D_3$ . So in the latter case, in grounded semantics  $r$  is neither sceptically nor credulously justified. However, in preferred and stable semantics we then obtain two alternative extensions: the first contains  $D_4$  while the second instead contains  $B_2$  and  $B_3$  and so excludes  $A_2$  and  $A_3$ . So in the latter case  $r$  is credulously, but not sceptically justified under stable and preferred semantics.

### 3.5. More on argument orderings

A well studied use of preferences in the non-monotonic logic literature is based on the use of priority orderings over formulae in the language or defeasible inference rules. If  $ASPIC^+$  is to be used as a framework for giving argumentation-based characterizations of non-monotonic formalisms augmented with priorities, then it needs to provide an account of how these priority orderings can be ‘lifted’ to preferences over arguments. Now the first thing to note is that if your use of  $ASPIC^+$  involves using defeasible inference rules and ordinary premises, then both may come equipped with priority orderings  $\leq$  on  $\mathcal{R}_d$  and  $\leq'$  on  $\mathcal{K}_p$ . We assume that these priority orderings are distinct to allow for the ontological nature of the rules and premises to be distinct. For example, the ordinary premises may represent the content of percepts from sensors or of witness testimonies, whose priority ordering reflects the relative reliability of the sensors, respectively witnesses. The defeasible rules may, for example, be prioritized based on probabilistic strength, on temporal precedence (defeasible rules acquired later are preferred to those acquired earlier), on the basis of principles of legal precedence, and so on. The challenge is to then define a preference over arguments  $A$  and  $B$  based on the priorities over their constituent ordinary premises *and* defeasible rules.

We now define two argument preference orderings, called the weakest-link and last-link orderings. These orderings are in turn based on priority orderings  $\leq$  on  $\mathcal{R}_d$  and  $\leq'$  on  $\mathcal{K}_p$ , where as usual,  $X <^{(\prime)} Y$  iff  $X \leq^{(\prime)} Y$  and  $Y \not\leq^{(\prime)} X$  (note that we may represent orderings in terms of the strict counterpart they define). However, these priorities relate individual defeasible rules, respectively ordinary premises, whereas when comparing two arguments, we want to compare them on the (possibly non-singleton) *sets* of rules/premises that these arguments are constructed from. So, to define these argument preferences, we need to first define a set ordering  $\trianglelefteq_s$  over sets of rules/premises (where  $\Gamma \trianglelefteq_s \Gamma'$  iff  $\Gamma \leq_s \Gamma'$  and  $\Gamma \not\leq_s \Gamma'$ )

---

determining conflict freeness and acceptability of arguments.

Note that for any sets of defeasible rules/ordinary premises  $\Gamma$  and  $\Gamma'$ , we intuitively want that:

- 1) if  $S$  is the empty set, it cannot be that  $S \preceq_s S'$  (hence it cannot be that  $S \triangleleft_s S'$ );
- 2) if  $S'$  is the empty set, it must be that  $S \preceq_s S'$  (and hence, given 1),  $S \triangleleft_s S'$  for any non-empty  $S$ .

In other words, arguments that have no defeasible rules (ordinary premises) are, modulo the premises (rules), strictly stronger than (preferred to) arguments that have defeasible rules (ordinary premises). Hence the following definition explicitly imposes these constraints, and then gives two alternative ways of defining  $\preceq_s$ ; the so called Elitist and Democratic ways (i.e.,  $s = \text{Eli}$  and  $\text{Dem}$  respectively).  $\text{Eli}$  compares sets on their minimal and  $\text{Dem}$  on their maximal elements.

**Definition 3.19:** [Orderings  $\preceq_s$ ] Let  $\Gamma$  and  $\Gamma'$  be finite sets<sup>1</sup>. Then  $\preceq_s$  is defined as follows:

- (1) If  $\Gamma = \emptyset$  then  $\Gamma \not\preceq_s \Gamma'$ ;
- (2) If  $\Gamma' = \emptyset$  and  $\Gamma \neq \emptyset$  then  $\Gamma \preceq_s \Gamma'$ ;  
else, assuming a preordering  $\leq$  over the elements in  $\Gamma \cup \Gamma'$ , then if:
- (3)  $s = \text{Eli}$ :  
 $\Gamma \preceq_{\text{Eli}} \Gamma'$  if  $\exists X \in \Gamma$  s.t.  $\forall Y \in \Gamma', X \leq Y$ .  
else, if:
- (4)  $s = \text{Dem}$ :  
 $\Gamma \preceq_{\text{Dem}} \Gamma'$  if  $\forall X \in \Gamma, \exists Y \in \Gamma', X \leq Y$ .

Note that one can directly define  $\Gamma \triangleleft_s \Gamma'$  in terms of the strict counter-part  $<$  of  $\leq$ , replacing  $\preceq_s$  by  $\triangleleft_s$  and  $\leq$  by  $<$  in (3) and (4) of the above definition.

Henceforth, we will assume that  $\preceq_{\text{Eli}}$  is used to compare sets of rules/premises.

Now the **last-link principle** prefers an argument  $A$  over another argument  $B$  if the last defeasible rules used in  $B$  are less preferred ( $\preceq_s$ ) than the last defeasible rules in  $A$  or, in case both arguments are strict, if the premises of  $B$  are less preferred than the premises of  $A$ . The concept of ‘last defeasible rules’ is defined as follows.

**Definition 3.20:** [Last defeasible rules] Let  $A$  be an argument.

- $\text{LastDefRules}(A) = \emptyset$  iff  $\text{DefRules}(A) = \emptyset$ .
- If  $A = A_1, \dots, A_n \Rightarrow \phi$ , then  $\text{LastDefRules}(A) = \{\text{Conc}(A_1), \dots, \text{Conc}(A_n) \Rightarrow \phi\}$ ,  
else  $\text{LastDefRules}(A) = \text{LastDefRules}(A_1) \cup \dots \cup \text{LastDefRules}(A_n)$ .

A simple example with more than one last defeasible rule is with  $\mathcal{K} = \{p; q\}$ ,  $\mathcal{R}_s = \{r, s \rightarrow t\}$  and  $\mathcal{R}_d = \{p \Rightarrow r; q \Rightarrow s\}$ . Then for the argument  $A$  for  $t$  we have that  $\text{LastDefRules}(A) = \{p \Rightarrow r; q \Rightarrow s\}$ .

The above definition is now used to compare pairs of arguments as follows:

**Definition 3.21:** [Last link principle] Let  $A$  and  $B$  be two arguments. Then  $A \preceq B$  iff:

- (1)  $\text{LastDefRules}(A) \preceq_s \text{LastDefRules}(B)$ ; or
- (2)  $\text{LastDefRules}(A)$  and  $\text{LastDefRules}(B)$  are empty and  $\text{Prem}(A) \preceq_s \text{Prem}(B)$ .

Note that although  $A \prec B$  holds in the case that  $A \preceq B$  and  $B \not\preceq A$  according to Definition 3.21, one can directly define  $A \prec B$  in terms of  $\triangleleft_s$ , by replacing  $\preceq$  by  $\prec$  and  $\preceq_s$  by  $\triangleleft_s$  in the above definition.

**Example 3.22** Suppose in our running example that  $u \prec' s$ ,  $x \prec' s$ ,  $d_2 < d_5$  and  $d_4 < d_2$ . Applying the last-link ordering, we must, to check whether  $C_3$  defeats  $B_1$ ,

<sup>1</sup>Notice that it suffices to restrict  $\triangleleft$  to finite sets since *ASPIC*<sup>+</sup> arguments are assumed to be finite (in Definition 3.14) and so their sets of ordinary premises/defeasible rules must be finite.

compare  $\text{LastDefRules}(C_3) = \{d_4\}$  with  $\text{LastDefRules}(B_1) = \emptyset$ . Clearly,  $\{d_4\} \triangleleft_{\text{Eli}} \emptyset$ , so  $C_3 \prec B_1$ , so  $C_3$  does not defeat  $B_1$ . Next, to check the conflict between  $B_2$  and  $D_4$  we compare  $\text{LastDefRules}(B_2) = \{d_2\}$  with  $\text{LastDefRules}(D_4) = \{d_5\}$ . Since  $d_2 < d_5$  we have that  $\text{LastDefRules}(B_2) \triangleleft_{\text{Eli}} \text{LastDefRules}(D_4)$ , so  $D_4$  strictly defeats  $B_2$ .

The **weakest-link principle** considers not the last but all uncertain elements in an argument. In the following definition,  $\text{Prem}_p(A) = \text{Prem}(A) \cap \mathcal{K}_p$ .

**Definition 3.23:** [Weakest link principle] Let  $A$  and  $B$  be two arguments. Then  $A \preceq B$  iff

- (1) If both  $B$  and  $A$  are strict, then  $\text{Prem}_p(A) \preceq_s \text{Prem}_p(B)$ , else;
- (2) If both  $B$  and  $A$  are firm, then  $\text{DefRules}(A) \preceq_s \text{DefRules}(B)$ , else;
- (3)  $\text{Prem}_p(A) \preceq_s \text{Prem}_p(B)$  and  $\text{DefRules}(A) \preceq_s \text{DefRules}(B)$

Note that in same way as in Definition 3.21 one can directly define  $A \prec B$  under the weakest link principle, in terms of  $\triangleleft_s$ .

**Example 3.24** If in our running example we apply the weakest-link ordering, then we must, to check whether  $C_3$  defeats  $B_1$ , first compare  $\text{Prem}_p(C_3) = \{u\}$  with  $\text{Prem}_p(B_1) = \{s\}$ . Since  $u <' s$  we have that  $\text{Prem}_p(C_3) \triangleleft_{\text{Eli}} \text{Prem}_p(B_1)$ . Then we must compare  $\text{DefRules}(C_3) = \{d_4\}$  with  $\text{DefRules}(B_1) = \emptyset$ . We have as above that  $\{d_4\} \triangleleft_{\text{Eli}} \emptyset$ . So  $C_3 \prec B_1$  and so  $C_3$  does not defeat  $B_1$ . Next, to check the conflict between  $B_2$  and  $D_4$  we must first compare  $\text{Prem}_p(B_2) = \{s\}$  with  $\text{Prem}_p(D_4) = \{u, x\}$ . Since both  $u <' s$  and  $x < s'$  we have that  $\text{Prem}_p(D_4) \triangleleft_{\text{Eli}} \text{Prem}_p(B_2)$ . We must then compare  $\text{DefRules}(B_2) = \{d_2\}$  with  $\text{DefRules}(D_4) = \{d_4, d_5\}$ . Since  $d_4 < d_2$  we now have that  $\text{DefRules}(D_4) \triangleleft_{\text{Eli}} \text{DefRules}(B_2)$ . So  $D_4 \prec B_2$  and  $B_2$  strictly defeats  $D_4$ .

We next discuss with two examples when the last-, respectively, weakest-link ordering may be more suitable. Consider first the following example on whether people misbehaving in a university library may be denied access to the library.<sup>1</sup>

**Example 3.25** Let  $\mathcal{K}_p = \{\text{Snores}; \text{Professor}\}$ ,  $\mathcal{R}_d =$

$$\begin{aligned} &\{\text{Snores} \Rightarrow_{d_1} \text{Misbehaves}; \\ &\text{Misbehaves} \Rightarrow_{d_2} \text{AccessDenied}; \\ &\text{Professor} \Rightarrow_{d_3} \neg \text{AccessDenied}\}. \end{aligned}$$

Assume that  $\text{Snores} <' \text{Professor}$  and  $d_1 < d_2$ ,  $d_1 < d_3$ ,  $d_3 < d_2$ , and consider the following arguments.

$$\begin{array}{ll} A_1: & \text{Snores} & B_1: & \text{Professor} \\ A_2: & A_1 \Rightarrow \text{Misbehaves} & B_2: & B_1 \Rightarrow \neg \text{AccessDenied} \\ A_3: & A_2 \Rightarrow \text{AccessDenied} & & \end{array}$$

Let us apply the ordering to the arguments  $A_3$  and  $B_2$ . The rule sets to be compared are  $\text{LastDefRules}(A_3) = \{d_2\}$  and  $\text{LastDefRules}(B_2) = \{d_3\}$ . Since  $d_3 < d_2$  we have that  $\text{LastDefRules}(B_2) \triangleleft_{\text{Eli}} \text{LastDefRules}(A_3)$ , hence  $B_2 \prec A_3$ . So  $A_3$  strictly defeats  $B_2$  (i.e.,  $A_3$  defeats  $B_2$  but  $B_2$  does not defeat  $A_3$ ). We therefore have that  $A_3$  is justified in any semantics, so we conclude *AccessDenied*.

With the weakest-link principle the ordering between  $A_3$  and  $B_2$  is different. Both  $A$  and  $B$  are plausible and defeasible so we are in case (3) of Definition 3.23. Since  $\text{Snores} <' \text{Professor}$ , we have that  $\text{Prem}_p(A_3) \triangleleft_{\text{Eli}} \text{Prem}_p(B_2)$ . Furthermore, the rule sets to be compared are now  $\text{DefRules}(A_3) = \{d_1, d_2\}$  and  $\text{DefRules}(B_2) = \{d_3\}$ . Since  $d_1 < d_3$  we have that  $\text{DefRules}(A_3) \triangleleft_{\text{Eli}} \text{DefRules}(B_2)$ . So now we have that  $A_3 \prec B_2$ . Hence  $B_2$  now strictly defeats  $A_3$  and we conclude instead that  $\neg \text{AccessDenied}$ .

<sup>1</sup>In all examples below, sets that are not specified are assumed to be empty.

Which outcome in this example is better? Some have argued that the last-link ordering gives the better outcome since the conflict really is between the two legal rules about whether someone may be denied access to the library, while  $d_1$  just provides a sufficient condition for when a person can be said to misbehave. The existence of a conflict on whether someone may be denied access to the library is in no way relevant for the issue of whether a person misbehaves when snoring. More generally, it has been argued that for reasoning with legal (and other normative) rules the last-link ordering is appropriate.

However, an example can be given of exactly the same form but with the legal rules replaced by empirical generalisations, and in that case intuitions seem to favour the weakest-link ordering:

**Example 3.26** Let  $\mathcal{K}_p = \{BornInScotland; FitnessLover\}$ ,  $\mathcal{R}_d = \{BornInScotland \Rightarrow_{d_1} Scottish; Scottish \Rightarrow_{d_2} LikesWhisky; FitnessLover \Rightarrow_{d_3} \neg LikesWhisky\}$ .

Assume that  $BornInScotland <' FitnessLover$  and  $d_1 < d_2$ ,  $d_1 < d_3$ ,  $d_3 < d_2$ , and consider the following arguments.

$A_1: BornInScotland$	$B_1: FitnessLover$
$A_2: A_1 \Rightarrow Scottish$	$B_2: B_1 \Rightarrow \neg LikesWhisky$
$A_3: A_2 \Rightarrow LikesWhisky$	

This time it seems reasonable to conclude  $\neg LikesWhisky$ , since the epistemic uncertainty of the premise and  $d_1$  of  $A_3$  should propagate to weaken  $A_3$ . And this is the outcome given by the weakest-link ordering. So it could be argued that for epistemic reasoning the weakest-link ordering is appropriate.

#### 4. Ways to use the framework

As should be clear by now,  $ASPIC^+$  is not a system but a framework for specifying systems.  $ASPIC^+$  leaves you fully free to make choices as to the logical language, the strict and defeasible inference rules, the axioms and ordinary premises in your knowledge base, and the argument preference ordering. In this section we discuss various more or less principled ways to make your choices, and then show specific uses of  $ASPIC^+$ .

##### 4.1. Choosing strict rules, axioms and defeasible rules

###### 4.1.1. Domain specific strict inference rules

When designing your  $ASPIC^+$  system, you can specify domain specific strict inference rules, as illustrated by the following example (based on Example 4 of Caminada and Amgoud 2007) in which the strict inference rules capture definitional knowledge, namely, that bachelors are not married.<sup>1</sup>

**Example 4.1** Let  $\mathcal{R}_d = \{d_1, d_2\}$  and  $\mathcal{R}_s = \{s_1, s_2\}$ , where:

$d_1 = WearsRing \Rightarrow Married$	$s_1 = Married \rightarrow \neg Bachelor$
$d_2 = PartyAnimal \Rightarrow Bachelor$	$s_2 = Bachelor \rightarrow \neg Married$

<sup>1</sup>In the examples that follow we may use terms of the form  $s_i$ ,  $d_i$  or  $f_i$ , to identify strict or defeasible inference rules or items from the knowledge base. We will assume that the  $d_i$  names are those assigned by the  $n$  function of Definition 3.2; sometimes we will attach these names to the  $\Rightarrow$  symbol. Note that the  $s_i$  and  $f_i$  names have no formal meaning and are for ease of reference only.

Finally, let  $\mathcal{K}_p = \{WearsRing, PartyAnimal\}$ . Consider the following arguments.

$A_1$ :	$WearsRing$	$B_1$ :	$PartyAnimal$
$A_2$ :	$A_1 \Rightarrow Married$	$B_2$ :	$B_1 \Rightarrow Bachelor$
$A_3$ :	$A_2 \rightarrow \neg Bachelor$	$B_3$ :	$B_2 \rightarrow \neg Married$

We have that  $A_3$  rebuts  $B_3$  on its subargument  $B_2$  while  $B_3$  rebuts  $A_3$  on its subargument  $A_2$ . Note that  $A_2$  does not rebut  $B_3$ , since  $B_3$  applies a strict rule; likewise for  $B_2$  and  $A_3$ .

Notice that in the above example, the rules  $s_1$  and  $s_2$  are ‘transpositions’ of each other, and  $\mathcal{R}_s$  is ‘closed under transposition’, in the sense that:

**Definition 4.2:** Let  $AT = (AS, \mathcal{K})$  be an argumentation theory. Then  $AT$  is *closed under transposition* iff if  $\phi_1, \dots, \phi_n \rightarrow \psi \in \mathcal{R}_s$ , then for  $i = 1 \dots n$ ,  $\phi_1, \dots, \phi_{i-1}, \neg\psi, \phi_{i+1}, \dots, \phi_n \rightarrow \neg\phi_i \in \mathcal{R}_s$ .

In general it is a good idea to ensure that your theory is closed under transposition. Proponents of this idea argue that this follows from the intuitive meaning of a strict rule as capturing deductive, that is, perfect inference: a strict rule  $q \rightarrow \neg s$  expresses that if  $q$  is true, then this guarantees the truth of  $\neg s$ , no matter what. Hence, if we have  $s$ , then  $q$  cannot hold, otherwise we would have  $\neg s$ . In general, if the negation of the consequent of a strict rule holds, then we cannot have all its antecedents, since if we had all of them, then its consequent would hold. This is the very meaning of a strict rule. So it is very reasonable to include in  $\mathcal{R}_s$  the transposition of a strict rule that is in  $\mathcal{R}_s$ . A second reason for ensuring closure under transposition is that it ensures satisfaction of Caminada and Amgoud (2007)’s rationality postulates, as illustrated later in Section 4.2.

#### 4.1.2. Strict inference rules and axioms based on deductive logics

Some find the use of domain-specific strict inference rules rather odd: why not instead express them as material implications in  $\mathcal{L}$  and put them in the knowledge base as axiom premises? These people want to reserve the strict inference rules for general patterns of deductive inference, since they say that this is what inference rules are meant for in logic. (Below we will see that the same issue arises with regard to the choice of defeasible rules, but we ignore that issue for the moment).  $ASPIC^+$  allows you to do this by basing your strict inference rules (and axioms) on a deductive logic of your choice. You can do so by choosing a semantics for your choice of  $\mathcal{L}$  with an associated monotonic notion of semantic consequence, and then filling  $\mathcal{R}_s$  with rules that are sound with respect to that semantics. For example, suppose you want it to conform to classical logic: you want to choose a standard propositional (or first-order) language, and you want that arguments can contain any classically valid inference step over this language. In  $ASPIC^+$  you can achieve this in two ways, a crude way and a sophisticated way.

A crude way is to simply put all valid propositional (or first-order) inferences over your language of choice in  $\mathcal{R}_s$ . So if you have chosen a propositional language, then you define the content of  $\mathcal{R}_s$  as follows. (where  $\vdash_{PL}$  denotes standard propositional-logic consequence). For any finite  $S \subseteq \mathcal{L}$  and any  $\varphi \in \mathcal{L}$ :<sup>1</sup>

$$S \rightarrow \varphi \in \mathcal{R}_s \text{ if and only if } S \vdash_{PL} \varphi$$

In fact, with this choice of  $\mathcal{R}_s$ , strict parts of an argument don’t need to be more than one step long. For example, if rules  $S \rightarrow \varphi$  and  $\varphi \rightarrow \psi$  are in  $\mathcal{R}_s$ , then  $S \cup \{\varphi\} \rightarrow \psi$  will also be in  $\mathcal{R}_s$ . Note also that using this method your strict rules will be closed under transposition, because of the properties of classical logic. The proof is easy: suppose  $p \rightarrow q$  is in  $\mathcal{R}_s$  for some  $p$  and  $q$ . Then we know that  $p \vdash_{PL} q$ , so (by the deduction theorem

<sup>1</sup>Although antecedents of rules formally are sequences of formulas, we will sometimes abuse notation and write them as sets.

for classical logic)  $\vdash_{PL} p \supset q$  so (by the properties of  $\vdash_{PL}$ ) we have  $\vdash_{PL} -q \supset -p$  so (by the other half of the deduction theorem) we have  $-q \vdash_{PL} -p$ , so (by choice of  $\mathcal{R}_s$ )  $-q \rightarrow -p \in \mathcal{R}_s$ .

Let us illustrate the crude approach with a variation on Example 4.1. We retain the defeasible rules  $d_1$  and  $d_2$  but we replace the domain-specific strict rules  $s_1$  and  $s_2$  with a single material implication  $Married \supset \neg Bachelor$  in  $\mathcal{K}_n$ . Moreover, we put all propositionally valid inferences over our language in  $\mathcal{R}_s$ . Then the arguments change as follows:

$A_1: \text{WearsRing}$ $A_2: A_1 \Rightarrow Married$ $A_3: Married \supset \neg Bachelor$ $A_4: A_2, A_3 \rightarrow \neg Bachelor$	$B_1: \text{PartyAnimal}$ $B_2: B_1 \Rightarrow Bachelor$ $B_3: Married \supset \neg Bachelor$ $B_4: B_2, B_3 \rightarrow \neg Married$
--	--

Now  $A_4$  rebuts  $B_4$  on  $B_2$  while  $B_4$  rebuts  $A_4$  on  $A_2$ .

A sophisticated way to base the strict part of  $ASPIC^+$  on a deductive logic of your choice is to build an existing axiomatic system for your logic into  $ASPIC^+$ . You can include its axiom(s) (typically a handfull) in  $\mathcal{K}_n$  and its inference rule(s) (typically just one or a few) in  $\mathcal{R}_s$ . For example, there are axiomatic systems for classical logic with just four axioms and just one inference rule, namely, modus ponens (i.e.  $\varphi \supset \psi, \varphi \rightarrow \psi$ )<sup>1</sup>. With this choice of  $\mathcal{R}_s$  strict parts of an argument could be very long, since in logical axiomatic systems proofs of even trivial validities might be long. However, this difference with the crude way is not very big, since if we want to be crude, we must, to know whether  $S \rightarrow \varphi$  is in  $\mathcal{R}_s$ , first construct a propositional proof of  $\varphi$  from  $S$ .

With the sophisticated way of building classical logic into our argumentation system, argument  $A_4$  in our example stays the same, since modus ponens is in  $\mathcal{R}_s$ . However, argument  $B_4$  will change, since modus tollens is not in  $\mathcal{R}_s$ . In fact,  $B_4$  will be replaced by a sequence of strict rule applications, together being an axiomatic proof of  $\neg Married$  from  $Married \supset \neg Bachelor$  and  $Bachelor$ .

Which approach is more natural? We think that the crude way is more like how people reason: people often summarise chunks of deductive reasoning in one step. But if you want to implement such reasoning on a computer, then the crude and sophisticated way do not differ much.

However, note that in the sophisticated method, closure under transposition may not hold; our example above does not contain modus tollens (that is,  $\varphi \supset \psi, -\psi \rightarrow -\varphi$ ). But we have already argued that the contrapositive reasoning yielded by the inclusion of transpositions is a desirable feature. Is this a problem for this method? No, since this reasoning can also be enforced without explicitly requiring transpositions of rules. Recall that  $S \vdash \varphi$  was defined as ‘there exists a strict argument for  $\varphi$  with all premises taken from  $S$ ’. Now it turns out that if  $\vdash$  contraposes, then this is just as good as closure of the strict rules under transposition. Contraposition of  $\vdash$  means that if  $S \vdash \varphi$ , then if we replace one element  $s$  of  $S$  with  $-\varphi$ , then  $-s$  is strictly implied. Now the point is that if  $\vdash$  corresponds to classical provability (as we have made it by our choice of axioms and inference rules), then  $\vdash$  does indeed contrapose.

**Definition 4.3:** Let  $AT = (AS, \mathcal{K})$  be an argumentation theory. We say that  $AT$  is *closed under contraposition* iff for all  $S \subseteq \mathcal{L}$ ,  $s \in S$  and  $\phi$ , if  $S \vdash \phi$ , then  $S \setminus \{s\} \cup \{-\phi\} \vdash -s$ .

Again, as will be discussed in Section 4.2, closure under contraposition also ensures satisfaction of rationality postulates.

We end this section by stating a quite general result on a class of logics that, if embedded in  $ASPIC^+$ , ensures closure of the strict rules under contraposition. In Amgoud and Besnard (2009) the idea was introduced to base argumentation logics on so-called

---

<sup>1</sup>As explained above, this strictly speaking is not a rule but a scheme or rules, with meta variables ranging over  $\mathcal{L}$ .



Tarskian abstract logics. Very briefly, abstract logics assume just some unspecified logical language  $\mathcal{L}$  and a consequence operator over this language, which to each subset of  $\mathcal{L}$  assigns a subset of  $\mathcal{L}$  (its logical consequences). Tarski then assumed a number of constraints on  $Cn$ , which we need not repeat here. Finally, Tarski defined a set  $S \subseteq \mathcal{L}$  as *consistent* iff  $Cn(S) \neq \mathcal{L}$ .

Now Amgoud and Besnard (2009)’s idea was to define an argument as a pair  $(S, p)$  where  $S \subseteq \mathcal{L}$  and  $p \in \mathcal{L}$ , where  $S$  is consistent,  $p \in Cn(S)$  and  $S$  is minimal in satisfying all these conditions. In  $ASPIC^+$  Tarski’s notion of an abstract logic can be used to generate the strict rules, via the following constraint (for any finite  $S$ ):

$$S \rightarrow p \in \mathcal{R}_s \text{ iff } p \in Cn(S)$$

It turns out that any AT with this choice of strict rules satisfies closure under contraposition. Strictly speaking, this only holds under some assumptions on the relation between the  $Cn$  function and  $ASPIC^+$ ’s negation (note that Tarski did not make any assumption on the syntax of  $\mathcal{L}$ ), but these assumptions are quite natural. For the details we refer the reader to Section 5.2 of Modgil and Prakken (2013).

#### 4.1.3. Choosing defeasible inference rules

Let us return to the question of how to choose the defeasible rules. Can we derive them from a logic of our choice just as we can derive the strict rules from a logic of our choice if we want to? This is controversial. Some philosophers argue that all rule-like structures that we use in daily life are “inference licences” and so cannot be expressed in the logical object language. In this view, all that can be done is apply them to formulas from  $\mathcal{L}$  to support new formulas from  $\mathcal{L}$ . That is, these philosophers see all defeasible generalisations as inference rules, whether they are domain-specific or not.

Others (usually logicians) take a more standard-logic approach (e.g. Kraus et al. (1990), Pearl (1992)). They say that all contingent knowledge should be expressed in the object language, so they reject the idea of domain-specific defeasible inference rules (for the same reason they don’t like domain-specific strict rules). They would introduce a new connective into  $\mathcal{L}$ , let us write it as  $\rightsquigarrow$ , where they informally read  $p \rightsquigarrow q$  as something like “If  $p$  then normally/typically/usually  $q$ ”. They then want to give a model-theoretic semantics for this connective just as logicians give a model-theoretic semantics for all connectives. The main difference is that such semantics for defeasible conditionals do not look at *all* models of a theory to check whether it entails a formula (as semantics for deductive logics do) but only to a *preferred class* of models of the theory (for example, all models where things are as normal as possible). They would then add a strict inference rule  $S \rightarrow \varphi$  to  $\mathcal{R}_s$  just in case  $\varphi$  is true in *all* models of  $S$ , while they would add a defeasible inference rule  $S \Rightarrow \varphi$  to  $\mathcal{R}_d$  just in case  $\varphi$  is true in all *preferred* models of  $S$  but *not* in all models of  $S$ .

Now what inference rules for  $\rightsquigarrow$  could result from such an approach? On two things there is consensus between logicians: modus ponens for  $\rightsquigarrow$  is defeasibly but not deductively valid, so the rule  $\varphi \rightsquigarrow \psi, \varphi \Rightarrow \psi$  should go into  $\mathcal{R}_d$ . There is also consensus that contraposition for  $\rightsquigarrow$  is deductively invalid, so the rule  $\varphi \rightsquigarrow \psi \rightarrow \neg\psi \rightsquigarrow \neg\varphi$  should *not* go into  $\mathcal{R}_s$ . However, here the consensus ends. Should the defeasible analogue of this rule go into  $\mathcal{R}_d$  or not? Opinions differ at this point<sup>1</sup>.

Let us illustrate the difference between the two approaches with a further variation on Example 4.1. Above we used the approach where all defeasible generalisations are inference rules. We now replace the two domain-specific defeasible inference rules  $d_1$  and  $d_2$  with two object-level conditionals expressed in  $\mathcal{L}$  and now add them to  $\mathcal{K}_p$ :

<sup>1</sup>See Chapter 4 of Caminada (2004) for a very readable overview of the discussion.

$WearsRing \rightsquigarrow Married$   
 $PartyAnimal \rightsquigarrow Bachelor$

Moreover, we add defeasible modus ponens for  $\rightsquigarrow$  to  $\mathcal{R}_d$ :

$$\mathcal{R}_d = \{\varphi \rightsquigarrow \psi, \varphi \Rightarrow \psi\}$$

The arguments then change as follows (assuming the crude way of incorporating classical logic):

$A_1: WearsRing$	$B_1: PartyAnimal$
$A_2: WearsRing \rightsquigarrow Married$	$B_2: PartyAnimal \rightsquigarrow Bachelor$
$A_3: A_1, A_2 \Rightarrow Married$	$B_3: B_1, B_2 \Rightarrow Bachelor$
$A_4: Married \supset \neg Bachelor$	$B_4: Married \supset \neg Bachelor$
$A_5: A_3, A_4 \rightarrow \neg Bachelor$	$B_5: B_3, B_4 \rightarrow \neg Married$

Now  $A_5$  rebuts  $B_5$  on  $B_3$  while  $B_5$  rebuts  $A_5$  on  $A_3$ .

Concluding, if you want, you can base at least some of your choices concerning defeasible inference rules on model-theoretic semantics for nonmonotonic logics. However, it is an open question whether a model-theoretic semantics is the *only* criterion by which we can choose our defeasible rules. Some have based their choice on other criteria, since they do not primarily see defeasible rules as logical inference rules but as principles of human cognition or rational action, so that they should be based on foundations other than semantics. For example, John Pollock based his defeasible reasons on his account of epistemology (the part of philosophy that studies how we can obtain knowledge). Others have based their choice of defeasible reasons on the study of argument schemes in informal argumentation theory. We give examples of both these approaches in Section 4.3.

#### 4.2. Satisfying rationality postulates

We are now in a position to state under what conditions  $ASPIC^+$  satisfies Caminada and Amgoud (2007)'s four rationality postulates. These are listed below (it is helpful to refer to concepts defined in Definition 3.4 when reading these postulates), adapted to the  $ASPIC^+$  framework. Let  $E$  be any complete extension of a Dung framework instantiated by  $ASPIC^+$  arguments and the  $ASPIC^+$  defeat relation (as defined in Section 3.4)

**Sub-argument Closure:** For any argument  $A$  in  $E$ , all sub-arguments of  $A$  are in  $E$ , i.e., for all  $A \in E$ : if  $A' \in \text{Sub}(A)$  then  $A' \in E$ .

**Closure under Strict Rules:** If  $E$  contains arguments with conclusions  $\alpha_1, \dots, \alpha_n$ , then any arguments obtained by applying only strict inference rules to these conclusions, are in  $E$ , i.e.,  $\{\text{Conc}(A) | A \in E\} = Cl(\{\text{Conc}(A) | A \in E\})$ .

**Direct Consistency:** The conclusions of arguments in  $E$  are directly consistent, i.e.,  $\{\text{Conc}(A) | A \in E\}$  is consistent.

**Indirect Consistency:** The conclusions of arguments in  $E$  are indirectly consistent, i.e.,  $Cl(\{\text{Conc}(A) | A \in E\})$  is consistent.

Depending on the choices outlined in Section 4.1.2, the first requirement for satisfying the above postulates is that your argumentation theory is closed under transposition or contraposition. This is because if neither property is satisfied, then since strict rule applications cannot be attacked, direct consistency may then be violated. This can be illustrated with the first version of Example 4.1. Suppose we only have one strict rule, namely,  $s_1$ . we cannot construct  $B_3$ , since  $B_3$  applies the now missing rule  $s_2$ . We still have that  $A_3$  rebuts  $B_2$ . Suppose now that  $d_1 < d_2$  and we apply the last-link argument ordering. Then  $A_3$  does not defeat  $B_2$ . In fact, no argument in the example is defeated, so we end up

with a single extension in all semantics, which contains arguments for both *Bachelor* and  $\neg$ *Bachelor* and so violates direct and indirect consistency.

However, with transposition this bad outcome is avoided: if we also have  $s_2$ , then argument  $B_3$  can be constructed, which rebuts  $A_3$  on  $A_2$ . Again applying the preference  $d_1 < d_2$  with the last-link ordering, we have that  $B_3$  strictly defeats  $A_2$ . Again we have a unique extension in all semantics, containing all arguments except  $A_2$  and  $A_3$ . This extension does not violate consistency.

**Example 4.4** Consider Example 3.7. As discussed in Example 3.18, if the argument ordering is such that  $C_3$  does not defeat  $B_1$ , then both arguments will be in the same extension, which thus violates consistency since the conclusions of these arguments contradict each other. However, if the transposition  $s \rightarrow \neg v$  of  $v \rightarrow \neg s$  is added to  $\mathcal{R}_s$ , then  $B_1$  can be continued to an argument for  $\neg v$ , which successfully rebuts  $C_3$  on  $C_2$ , excluding the consistency-violating extensions.

Some say that the above violation of consistency, before inclusion of the transposed rule, arises because  $ASPIC^+$  forbids attacks on strictly derived conclusions. Consistency would not be violated if  $B_2$  was allowed to attack  $A_3$  in the first version of Example 4.1. However, apart from the reasons discussed in Section 2, there is another reason for prohibiting attacks on strictly derived conclusions: if they are allowed, then extensions may not be strictly closed or indirectly consistent, even if the strict rules are closed under transposition. To see why, suppose we changed  $ASPIC^+$ 's definitions to allow attacks on strict conclusions, so that  $B_2$  attacks  $A_3$ ,  $A_2$  attacks  $B_3$ , and  $A_3$  and  $B_3$  attack each other in Example 4.1. Suppose also that all knowledge-base items and all defeasible rules in the example are of equal preference, and suppose we apply the weakest- or last-link argument ordering. Then all rebutting attacks in the example succeed. But then the set  $\{A_1, A_2, B_1, B_2\}$  is admissible and is in fact both a stable and preferred extension. But this violates the rationality postulates of strict closure and indirect consistency. The extension contains an argument for *Bachelor* but not for  $\neg$ *Married*, which strictly follows from it by rule  $s_2$ . Likewise, the extension contains an argument for *Married* but not for  $\neg$ *Bachelor*, which strictly follows from it by rule  $s_1$ . So the extension is not closed under strict rule application. Moreover, the extension is indirectly inconsistent, since its strict closure contains both *Married* and  $\neg$ *Married*, and both *Bachelor* and  $\neg$ *Bachelor*.

Other requirements for satisfying the postulates are that the axioms  $\mathcal{K}_n$  are indirectly consistent (axiom consistency) and the preference ordering is *reasonable*. The rationale for requiring the former is self-evident. A reasonable argument ordering essentially amounts to requiring that: 1) arguments that are both strict and firm are strictly preferred over all other arguments; 2) the strength (and implied relative preference) of an argument is determined exclusively by the defeasible rules and/or ordinary premises; 3) the preference ordering is acyclic, and if  $B \prec A$  then it must be that  $B' \prec A$  where  $B'$  is some maximal fallible (i.e., defeasible or plausible) sub-argument of  $B$  (for example in our running example  $C_2$  but not  $C_1$  is a maximal fallible argument of  $C_3$ ). We refer the reader to Modgil and Prakken (2013) for the technical definition of a reasonable ordering, suffice to say that in that paper it is shown that the weakest- and last-link argument orderings of Section 3.5 are reasonable.

We are now in a position to state an important result proved in Modgil and Prakken (2013) that if your argument theory is *well-defined*, in the sense that it satisfies axiom consistency, and transposition or contraposition, and your argument preference ordering is reasonable, then all four rationality postulates are satisfied by the  $ASPIC^+$  framework as defined in Section 3. Finally, note that if you do not include any strict rules or axiom premises in your argumentation theory, then the requirement that it be well defined obviously does not apply, but it is also worth noting that the preference ordering need *not* be reasonable in order that all four rationality postulates be satisfied (indeed no assumptions

as to the properties of the preference ordering are required in this case).

#### 4.3. Using ASPIC<sup>+</sup> to model argument schemes

We concluded Section 4.1.3 by remarking on the use of defeasible inference rules as principles of cognition in John Pollock's work and as argument schemes in informal argumentation theory. We now illustrate how both approaches can be formalised in ASPIC<sup>+</sup> and how strict inference rules can also be accommodated when doing so.

Let us first look in more detail at John Pollock's work. He formalised defeasible rules for reasoning patterns involving perception, memory, induction, temporal persistence and the statistical syllogism, as well as undercutters for these reasons.

In ASPIC<sup>+</sup> his principles of perception and memory can be written as follows:

$$\begin{aligned} d_p(x, \varphi): & \text{ Sees}(x, \varphi) \Rightarrow \varphi \\ d_m(x, \varphi): & \text{ Recalls}(x, \varphi) \Rightarrow \varphi \end{aligned}$$

In fact, these defeasible inference rules are schemes for all their ground instances (that is, for any instance where  $x$  and  $\varphi$  are replaced by ground terms denoting a specific perceiving agent and a specific perceived state of affairs). Therefore, their names  $d_p(x, \varphi)$  and  $d_m(x, \varphi)$  as assigned by the  $n$  function are in fact also schemes for names. A proper name is obtained by instantiating these variables by the same ground terms as used to instantiate these variables in the scheme. Thus it becomes possible to formulate undercutters for one instance of the scheme (say for Jan who saw John in Amsterdam) while leaving another instance unattacked (say for Bob who saw John in Holland Park). Note, finally, that these schemes assume a naming convention for formulas in a first-order language, since  $\varphi$  is a term in the antecedent while it is a well-formed formula in the consequent. In the remainder we will leave this naming convention implicit.

Now undercutters for  $d_p$  state circumstances in which perceptions are unreliable, while undercutters of  $d_m$  state conditions under which memories may be flawed. For example, a well-known cause of false memories of events is that the memory is distorted by, for instance, seeing pictures in the newspaper or watching a TV programme about the remembered event. A general undercutter for distorted memories could be

$$u_m(x, \varphi): \text{ DistortedMemory}(x, \varphi) \Rightarrow \neg d_m(x, \varphi)$$

combined with information such as

$$\forall x, \varphi (\text{SeesPicturesAbout}(x, \varphi) \supset \text{DistortedMemory}(x, \varphi))$$

Pollock's epistemic inference schemes are in fact a subspecies of argument schemes. The notion of an argument scheme was developed in philosophy and is currently an important topic in the computational study of argumentation. Argument schemes are stereotypical non-deductive patterns of reasoning, consisting of a set of premises and a conclusion that is presumed to follow from them. Uses of argument schemes are evaluated in terms of critical questions specific to the scheme. An example of an epistemic argument scheme is the scheme from the position to know (Walton 1996, pp. 61–63):

$$\frac{\begin{array}{l} A \text{ is in the position to know whether } P \text{ is true} \\ A \text{ asserts that } P \text{ is true} \end{array}}{P \text{ is true}}$$

Walton gives this scheme three critical questions:

1. Is  $A$  in the position to know whether  $P$  is true?
2. Did  $A$  assert that  $P$  is true?
3. Is  $A$  an honest (trustworthy, reliable) source?

A natural way to formalise reasoning with argument schemes is to regard them as defeasible inference rules and to regard critical questions as pointers to counterarguments. For example, in the scheme from the position to know questions (1) and (2) point to underminers (of, respectively, the first and second premise) while questions (3) points to undercutters (the exception that the person is for some reason not credible).

Accordingly, we formalise the position to know scheme and its undercutter as follows:

$$\begin{aligned} d_w(x, \varphi): & \text{ PositionToKnow}(x, \varphi), \text{ Says}(x, \varphi) \Rightarrow \varphi \\ u_w(x, \varphi): & \neg\text{Credible}(x) \Rightarrow \neg d_w(x, \varphi) \end{aligned}$$

We will now illustrate the modelling of both Pollock’s defeasible reasons and Walton’s argument schemes with our example from Section 2, focusing on a specific class of persons who are in the position to know, namely, witnesses. In fact, witnesses always report about what they observed in the past, so they will say something like “I remember that I saw that John was in Holland Park”. Thus an appeal to a witness testimony involves the use of three schemes: first the position to know scheme is used to infer that the witness indeed remembers that he saw that John was in Holland Park, then the memory scheme is used to infer that he indeed saw that John was in Holland Park, and finally, the perception scheme is used to infer that John was indeed in Holland Park. Now recall that John was a suspect in a robbery in Holland Park and that Jan testified that he saw John in Amsterdam on the same morning, while Jan is a friend of John. Suppose now we also receive information that Bob read newspaper reports about the robbery in which a picture of John was shown. One way to model this in  $ASPIC^+$  is as follows.

The knowledge base consists of the following facts (since we don’t want to dispute them, we put them in  $\mathcal{K}_n$ ):

- $f_1$ :  $\text{PositionToKnow}(\text{Bob}, \text{Recalls}(\text{Bob}, \text{Sees}(\text{Bob}, \text{InHollandPark}(\text{John}))))$
- $f_2$ :  $\text{Says}(\text{Bob}, \text{Recalls}(\text{Bob}, \text{Sees}(\text{Bob}, \text{InHollandPark}(\text{John}))))$
- $f_3$ :  $\text{SeesPicturesAbout}(\text{Bob}, \text{Sees}(\text{Bob}, \text{InHollandPark}(\text{John})))$
- $f_4$ :  $\forall x, \varphi. (\text{SeesPicturesAbout}(x, \varphi) \supset \text{DistortedMemory}(x, \varphi))$
- $f_5$ :  $\forall x. \text{InHollandPark}(x) \supset \text{InLondon}(x)$
- $f_6$ :  $\text{PositionToKnow}(\text{Jan}, \text{Recalls}(\text{Jan}, \text{Sees}(\text{Jan}, \text{InAmsterdam}(\text{John}))))$
- $f_7$ :  $\text{Says}(\text{Jan}, \text{Recalls}(\text{Jan}, \text{Sees}(\text{Jan}, \text{InAmsterdam}(\text{John}))))$
- $f_8$ :  $\text{Friends}(\text{Jan}, \text{John})$
- $f_9$ :  $\text{SuspectedRobber}(\text{John})$
- $f_{10}$ :  $\forall x, y, \varphi. \text{Friends}(x, y) \wedge \text{SuspectedRobber}(y) \wedge \text{InvolvedIn}(y, \varphi) \supset \neg\text{Credible}(x)$
- $f_{11}$ :  $\text{InvolvedIn}(\text{John}, \text{Recalls}(\text{Jan}, \text{Sees}(\text{Jan}, \text{InAmsterdam}(\text{John}))))$
- $f_{12}$ :  $\forall x. \neg(\text{InAmsterdam}(x) \wedge \text{InLondon}(x))$

Combining this with the schemes from perception, memory and position to know, we obtain the following arguments (for reasons of space we don’t list separate lines for arguments that just take an item from  $\mathcal{K}$ ).

- $A_3$ :  $f_1, f_2 \Rightarrow_{dw} \text{Recalls}(\text{Bob}, \text{Sees}(\text{Bob}, \text{InHollandPark}(\text{John})))$
- $A_4$ :  $A_3 \Rightarrow_{dm} \text{Sees}(\text{Bob}, \text{InHollandPark}(\text{John}))$
- $A_5$ :  $A_4 \Rightarrow_{dp} \text{InHollandPark}(\text{John})$
- $A_7$ :  $A_5, f_5 \rightarrow \text{InLondon}(\text{John})$

This argument is undercut (on  $A_4$ ) by the following argument applying the undercutter for the memory scheme:

$B_3: f_3, f_4 \rightarrow \text{DistortedMemory}(\text{Bob}, \text{Sees}(\text{Bob}, \text{InHollandPark}(\text{John})))$   
 $B_4: B_3 \Rightarrow_{um} \neg d_m(\text{Bob}, \text{Sees}(\text{Bob}, \text{InHollandPark}(\text{John})))$

Moreover,  $A_7$  is rebutted (on  $A_5$ ) by the following argument:

$C_3: f_6, f_7 \Rightarrow_{dw} \text{Recalls}(\text{Jan}, \text{Sees}(\text{Jan}, \text{InAmsterdam}(\text{John})))$   
 $C_4: C_3 \Rightarrow_{dm} \text{Sees}(\text{Jan}, \text{InAmsterdam}(\text{John}))$   
 $C_5: C_4 \Rightarrow_{dp} \text{InAmsterdam}(\text{John})$   
 $C_8: C_5, f_5, f_{12} \rightarrow \neg \text{InHollandPark}(\text{John})$

This argument is also undercut, namely, on  $C_3$  based on the undercutter of the position to know scheme:

$D_4: f_8, f_9, f_{10}, f_{11} \rightarrow \neg \text{Credible}(\text{Jan})$   
 $D_5: D_4 \Rightarrow_{uw} \neg d_w(\text{Jan}, \text{Recalls}(\text{Jan}, \text{Sees}(\text{Jan}, \text{InAmsterdam}(\text{John}))))$

Finally,  $C_8$  is rebutted on  $C_5$  by the following continuation of argument  $A_7$ :

$A_8: A_5, f_5, f_{12} \Rightarrow \neg \text{InAmsterdam}(\text{John})$

$A_8$  is in turn undercut by  $B_4$  (on  $A_4$ ) and rebutted by  $C_8$  (on  $A_5$ ).

Because of the two undercutting arguments, neither of the testimony arguments are credulously or sceptically justified in any semantics. Let us now see what happens if we do not have the two undercutters. Then we must apply preferences to the rebutting attack of  $C_8$  on  $A_5$  and to the rebutting attack of  $A_8$  on  $C_5$ . As it turns out, exactly the same preferences have to be applied in both cases, namely, those between the three defeasible-rule applications in the respective arguments. And this is what we intuitively want.

Finally, we note that counterarguments based on critical questions of argument schemes may themselves apply argument schemes. For example, we may believe that Jan and John are friends because another witness told our so. Or we may believe that Holland Park is in London because a London taxi driver told us so (an application of the so-called expert testimony scheme).

#### 4.4. Instantiations with no defeasible rules

All that has been said so far about ways to choose the strict rules applies irrespective of whether you also want to include defeasible rules in your argumentation system. In fact,  $ASPIC^+$  allows you to only use strict inference rules. Principled ways to do so are to base the strict rules on classical logic or indeed on any Tarskian consequence relation. In this way,  $ASPIC^+$  extends the classical-logic approach of Besnard and Hunter (2009) and the abstract-logic approach of Amgoud and Besnard (2009), by providing guidelines for using preferences to resolve inconsistencies in classical logic or any other underlying Tarskian logic. The use of preferences is of particular importance in such contexts, since in these contexts the stable and preferred extensions of Dung frameworks simply correspond to the maximal consistent subsets of the instantiating theories (Amgoud and Besnard 2013). One thus needs some ‘extra-logical’ means, such as preferences, to resolve inconsistencies.

The idea is as follows. Given a set  $S$  of wff in some language  $\mathcal{L}$  and a Tarskian consequence relation  $Cn$  over  $\mathcal{L}$  (note that classical consequence is such a Tarskian consequence relation), we let the axioms and defeasible inference rules be empty, and the strict rules defined as indicated in Section 4.1.2, namely, as  $S \rightarrow p \in \mathcal{R}_s$  iff  $p \in Cn(S)$ , for any finite  $S \subseteq \mathcal{L}$ . Furthermore, in keeping with the above mentioned classical, and more general Tarskian Logic approaches, we put an extra constraint on  $ASPIC^+$  arguments, namely, that their set of premises is indirectly consistent and, moreover, subset-minimal in applying their conclusion

For this special case all  $ASPIC^+$  arguments are strict, so all attacks are undermining attacks. In Modgil and Prakken (2013) it was shown that these  $ASPIC^+$  reconstructions of Tarskian and classical approaches are equivalent to the originals if these originals use a form of undermining attack. Moreover, the result stated in Section 4.1.2 – that any  $ASPIC^+$  AT with the strict rules derived from a Tarskian logic satisfies closure under contraposition — then implies that without preferences these reconstructions are well-defined and thus satisfy the rationality postulates<sup>1</sup>. Moreover, if these reconstructions are extended with a reasonable argument ordering, then this result also holds for the case with preferences. Thus the  $ASPIC^+$  framework has in fact been used to extend both the classical-logical approach of Besnard and Hunter (2009) and the more general Tarskian approach of Amgoud and Besnard (2009) with preferences in a way that satisfies all rationality postulates of Caminada and Amgoud (2007). A final result of Modgil and Prakken (2013) is that if a thus defined classical-logic instantiation of  $ASPIC^+$  is combined with a total priority ordering  $\leq'$ , then one obtains a correspondence with Brewka (1989)'s Preferred Subtheories.

#### 4.5. Illustrating uses of $ASPIC^+$ with and without defeasible rules

In this section we compare respective uses of  $ASPIC^+$  with and without defeasible rules in more detail. We first say more about the arguments of some that classical-logic simulations of defeasible rules may yield counterintuitive results. Let us assume a classical-logic instantiation of  $ASPIC^+$  as defined in Section 4.4 and formalise natural-language generalisations ‘If  $P$  then normally  $Q$ ’ as material implications  $P \supset Q$  put in  $\mathcal{K}_p$ . The idea is that since  $P \supset Q$  is an ordinary premise, its use as a premise can be undermined in exceptional cases. Observe that by classical reasoning we then have a strict argument for  $\neg Q \supset \neg P$ . Some say that this is problematic. Consider the following example: ‘Anyone who is a man usually has no beard’, so (strictly) ‘Anyone who has a beard usually is not a man’. This strikes some as counterintuitive, since we know that virtually everyone who has a beard is a man, so the contraposition of ‘If  $P$  then normally  $Q$ ’ cannot be deductively valid<sup>1</sup>.

A more refined classical approach is to give the material implication an extra normality condition  $N$ , which informally reads as ‘everything is normal as regards  $P$  implying  $Q$ ’, and which is also put in  $\mathcal{K}_p$ . The idea then is that exceptional cases give rise to underminers of  $N$ . However,  $(P \wedge N) \supset Q$  also deductively contraposes, namely, as  $(\neg Q \wedge N) \supset \neg P$ , so it seems that we still have the controversial deductive validity of contraposition for generalisations (in the beard and men example the contraposition of the rule with the added normality condition would read: ‘Anyone who has a beard and all is normal regarding men and having beards, usually is not a man’ !).

So far we only discussed reasons for belief but argumentation is often about what to do, prefer or value (what philosophers often call *practical reasoning*). Here too it has been argued on philosophical grounds that reasons for doing, preferring or valuing cannot be expressed in classical logic since they do not contrapose. This view can, of course, not be based on a statistical semantics for such reasons, since statistics only applies to reasoning about what is the case (what philosophers often call *epistemic reasoning*). Space limitations prevent us from giving more details about these philosophical arguments.

<sup>1</sup>In fact, given that the  $ASPIC^+$  arguments are restricted to those with consistent premises, satisfaction of the postulates also requires that if for some set of premises  $S$  is minimally (under set inclusion) indirectly inconsistent (see Definition 3.4), then  $\forall \phi \in S, S \setminus \{\phi\} \vdash \neg \phi$ . Modgil and Prakken (2013) show that this property is also satisfied for Tarskian based  $ASPIC^+$  ATs.

<sup>1</sup>One way to argue why classical simulations may give counter-intuitive results is to recall that a number of researchers provide statistical semantics for defeasible inference rules. These semantics regard a defeasible rule of the form  $P \Rightarrow Q$  as a qualitative approximation of the statement that the conditional probability of  $Q$ , given  $P$ , is high. The laws of probability theory then tell us that this does not entail that the conditional probability of  $\neg P$ , given  $\neg Q$ , is high. The problem with the classical-logic approach is then that it conflates this distinction by turning the conditional probability of  $Q$  given  $P$  into the unconditional probability of  $P \supset Q$ , which then has to be equal to the unconditional probability of  $\neg Q \supset \neg P$ .

We next illustrate two different ways to use *ASPIC*<sup>+</sup> with a detailed example. Both ways use classical logic in their strict part and use explicit preferences, but only the second way uses defeasible inference rules. The first way instead expresses defeasible generalisations as material implications with normality assumptions. The example will shed further light on the issue whether empirical generalisations can be represented in classical logic, and it will also motivate the use of axiom premises. Our example is a well-known one from the literature on nonmonotonic logic. Suppose a defeasible reasoner accepts all following natural-language statements are true. For the generalisations (1) and (2) this means that the reasoner accepts that they hold in general but that they may have exceptions.

- (1) Birds normally fly
- (2) Penguins normally don't fly
- (3) All penguins are birds
- (4) Penguins are abnormal birds with respect to flying
- (5) Tweety is a penguin

A defeasible reasoner then wants to know what can be concluded from this information about whether Tweety can fly. It seems uncontroversial to say that any defeasible reasoner will conclude that Tweety can fly.

We now formalise these statements with the just-explained method to represent empirical generalisations as material implications with explicit normality assumptions. We use a classical-logic instantiation of *ASPIC*<sup>+</sup> with preferences as defined above in Section 4.4.

- (1)  $bird \wedge \neg ab_1 \supset canfly$
- (2)  $penguin \wedge \neg ab_2 \supset \neg canfly$
- (3)  $penguin \supset bird$
- (4)  $penguin \supset ab_1$
- (5)  $penguin$

Let us first add these formulas to  $\mathcal{K}_p$ . The idea now is that the normality assumptions of a defeasible reasoner are expressed as additional statements  $\neg ab_1$  and  $\neg ab_2$ , also added to  $\mathcal{K}_p$ . We then define the preference ordering on  $\mathcal{K}_p$  such that all of (1-5) are strictly preferred over any of these two assumptions and that  $\neg ab_1 <' \neg ab_2$ .

We can then construct many arguments on the issue whether Tweety can fly. Note that  $\{1, 2, 3, 4, 5\} \cup \{\neg ab_1, \neg ab_2\}$  is minimally inconsistent, so if we take any single element out, the rest can be used to build an argument against it. This means that we can formally build arguments not just against the two normality assumptions but also against any of (1-5) (note the similarity with the fact that, as noted above, in classical-logic argumentation without preferences the stable and preferred extensions correspond to maximal consistent subsets of the knowledge base). With the weakest- or last-link ordering we do obtain the intuitive conclusion  $\neg canfly$ , but the fact that arguments against any of (1-5) can be built may be regarded as somewhat odd, since we just noted that a defeasible reasoner accepts (1-5) as given and is only interested in what follows from them.

Let us therefore move (1-5) to the axioms  $\mathcal{K}_n$ , so that they cannot be attacked. Then we have just a few arguments on the issue whether Tweety can fly: we have an argument  $\{1, 2, 3, 4, 5\} \cup \{\neg ab_2\} \rightarrow \neg canfly$ , which has one attacker, namely,  $\{1, 2, 3, 5\} \cup \{\neg ab_1\} \rightarrow ab_2$ . However, with the weakest- or last link principle this attacker does not defeat its target, since we have  $\neg ab_1 <' \neg ab_2$ . Hence  $\neg canfly$  is justified in any semantics. So at first sight it would seem that the classical-logic approach enriched with axiom premises adequately models reasoning with empirical generalisations.

However, this approach still has some things to explain, as can be illustrated by changing our example a little: above it was given as a matter of fact that Tweety is a penguin but in reality the particular 'facts' of a problem are not simply given but derived from information sources (sensors, testimonies, databases, the internet, and so on). Now in reality



none of these sources is fully reliable so inferring facts from them can only be done under the assumption that things are normal. So let us change the example by saying that Tweety was observed to be a penguin and that animals that are observed to be penguins *normally* are penguins. We change 5 to 5' and we add 6 to  $\mathcal{K}_n$ :

- (5')  $observed\_as\_penguin$
- (6)  $observed\_as\_penguin \wedge \neg ab_3 \supset penguin$

Moreover, we add  $\neg ab_3$  to  $\mathcal{K}_p$ . We can still build an argument that Tweety cannot fly, namely,  $\{1, 2, 3, 4, 5'\} \cup \{\neg ab_2, \neg ab_3\} \rightarrow \neg canfly$ . However, we can also build an attacker of this argument, namely  $\{1, 2, 3, 4, 5', 6\} \cup \{\neg ab_1, \neg ab_2\} \rightarrow ab_3$ . We can still obtain the intuitive outcome by preferring the assumption  $\neg ab_3$  over the assumption  $\neg ab_1$ . However, some have argued that this is an ad-hoc solution, since there would be no general principle on which such a preference can be based. The heart of the problem, they say, is the fact that the material implication satisfies contraposition, a property which, as we just mentioned, can be argued to be too strong for defeasible generalisations. In reality a defeasible reasoner would not even construct an argument against *penguin*. As can be easily checked, the same issues arise if we put (1-4,5',6) in  $\mathcal{K}_p$  while we then have our old issue back that arguments can be constructed against any element of  $\mathcal{K}_p$ .

Concluding so far, those who want to want to model 'default reasoning' in classical argumentation have to explain why arguments as the one for  $ab_3$  can be constructed and why it does not defeat the argument for  $\neg canfly$  (or alternatively, why the latter conclusion is not justified). Moreover, if they apply the first version of this approach, by putting all of  $\{1, 2, 3, 4, 5', 6\}$  in  $\mathcal{K}_p$ , then they also have to explain why arguments against any of these premises can be constructed and whether these arguments succeed as defeats.

Let us next formalise the example with domain-specific defeasible rules and with the strict rules still corresponding to classical logic.

- $d_1: bird \Rightarrow canfly$
- $d_2: penguin \Rightarrow \neg canfly$
- $d_3: observed\_as\_penguin \Rightarrow \neg penguin$
- $f_1: penguin \supset bird$
- $f_2: penguin \supset \neg r_1$
- $f_3: observed\_as\_penguin$

It now does not matter whether we put the facts in  $\mathcal{K}_n$  or  $\mathcal{K}_p$ , nor does it matter which priorities we define on  $\mathcal{K}_p$  or  $\mathcal{R}_d$ . We have the following arguments:

- $A_1: observed\_as\_penguin$
- $A_2: A_1 \Rightarrow penguin$
- $A_3: penguin \supset bird$
- $A_4: A_2, A_3 \Rightarrow canfly$
- $B_1: A_2 \Rightarrow \neg canfly$
- $C_1: A_2 \Rightarrow \neg r_1$

Note also that no argument can be built against the conclusion *penguin*. We have that  $A_4$  and  $B_1$  rebut each other while  $C_1$  undercuts  $A_4$ . Whatever the argument ordering between  $A_4$  and  $B_1$ , we thus obtain that the conclusion  $\neg canfly$  is justified in any semantics.

Concluding, the classical modelling of this example is simpler in that it only uses classical inference and does not have to rely on the notion of a defeasible inference rule. On the other hand, to obtain the intuitive outcome it needs more preferences than the modelling with defeasible rules, while the issue arises on which grounds these preferences can be stated. Moreover, if the classical approach regards all knowledge as fallible in principle, then it generates many more arguments than perhaps intuitively expected, at least many more than in the modelling with defeasible rules.

#### 4.6. Representing facts

$ASPIC^+$  allows you to represent facts in various ways, each with their pros and cons. *Disputable facts*  $\varphi$  can either be put as such in  $\mathcal{K}_p$  or as defeasible rules  $\Rightarrow \varphi$  with empty antecedents. An advantage of including disputable facts in  $\mathcal{K}_p$  is that thus  $ASPIC^+$  captures classical and abstract-logic argumentation with preferences as special cases. On the other hand, if disputable facts  $\varphi$  are represented as defeasible rules  $\Rightarrow \varphi$ , then the definition of the weakest- and last-link argument orderings becomes simpler, since then only sets of defeasible rules need to be compared. In addition, this choice removes the need for undermining attack, which simplifies the definitions of attack and defeat.

*Undisputable facts*  $\varphi$  can either be put as such in  $\mathcal{K}_n$  or as strict rules  $\rightarrow \varphi$  with empty antecedents. This choice does not make a difference for the weakest- or last-link argument ordering, since these orderings disregard axiom premises and strict rules. However, a disadvantage of representing undisputable fact  $\varphi$  as strict rules  $\rightarrow \varphi$  is that then the strict rules do not express a logic any more, so the above-mentioned theorems on definitions of  $\mathcal{R}_s$  in terms of Tarskian abstract logics do not apply any more.

#### 4.7. Summary

We have seen that  $ASPIC^+$  allows you to make any choice of axioms, strict and defeasible rules you like. You can choose domain-specific strict and/or defeasible inference rules, and you can choose logical strict and/or defeasible inference rules, for any deductive and/or nonmonotonic logic of your choice, good or bad. You can add logical axioms to  $\mathcal{K}_n$  but you can also add other information to  $\mathcal{K}_n$  that you don't want to put up for discussion, including even contingent facts. You can also base your defeasible rules on informal accounts of argument schemes. All that  $ASPIC^+$  tells you is how arguments can be built with your rules of choice, how they can be attacked, and how these attacks can be resolved, given an argument ordering of your choice. Moreover, we have some theorems about  $ASPIC^+$  that inform you about some properties of your choices.

### 5. Generalising negation in $ASPIC^+$

The notion of an argumentation system in Section 3.2, assumed a language  $\mathcal{L}$  closed under negation ( $\neg$ ), where the standard classical interpretation of  $\neg$  licenses a symmetric notion of conflict based attack, so that an argument consisting of an ordinary premise  $\phi$  or with a defeasible top rule concluding  $\phi$ , *symmetrically* attacks an argument consisting of an ordinary premise  $\neg\phi$  or with a defeasible top rule concluding  $\neg\phi$ . However, the  $ASPIC^+$  framework as presented in Modgil and Prakken (2013), Prakken (2010), accommodates a more general notion of conflict, by defining an argumentation system to additionally include a function  $\bar{\phantom{x}}$  that, for any wff  $\psi \in \mathcal{L}$ , specifies the set of wff's that are in conflict with  $\psi$ . With this idea, which is taken from assumption-based argumentation (Bondarenko et al. 1997, Dung et al. 2009), one can define both an asymmetric and symmetric notion of conflict-based attack. More formally:

**Definition 5.1:**  $\bar{\phantom{x}}$  is a function from  $\mathcal{L}$  to  $2^{\mathcal{L}}$ , such that:

- $\varphi$  is a *contrary* of  $\psi$  if  $\varphi \in \bar{\psi}$ ,  $\psi \notin \bar{\varphi}$ ;
- $\varphi$  is a *contradictory* of  $\psi$  (denoted by ' $\varphi = -\psi$ '), if  $\varphi \in \bar{\psi}$ ,  $\psi \in \bar{\varphi}$ ;
- each  $\varphi \in \mathcal{L}$  has at least one contradictory.

Note that classical negation is now a special case of the symmetric contradictory relation:  $\alpha \in \bar{\beta}$  iff  $\alpha$  is of the form  $\neg\beta$  or  $\beta$  is of the form  $\neg\alpha$  (i.e., for any wff  $\alpha$ ,  $\alpha$  and  $\neg\alpha$  are contradictories). Modgil and Prakken (2013) then redefine Definition 3.4's notion of

direct consistency so that a set  $S$  is *directly consistent* iff  $\nexists \psi, \varphi \in S$  such that  $\psi \in \overline{\varphi}$ . Also,  $\text{Conc}(A) \in \overline{\varphi}$  ( $\text{Conc}(A) \in \overline{n(r)}$ ) replaces  $\text{Conc}(A) = -\varphi$  ( $\text{Conc}(A) = -n(r)$ ) in Definition 3.10's definition of attacks.

With this, one can reconstruct assumption-based argumentation (ABA) in  $ASPIC^+$ , since as noted above,  $ABA$  also generalises the notion of conflict through the use of a  $\neg$  function. Indeed, this reconstruction is formally shown in Prakken (2010), in which assumption premises were distinguished from ordinary premises, and used to model ABA assumptions. However, one can do without such specialised premises, and model assumptions as ordinary premises. So, to summarise, an  $ASPIC^+$  reconstruction of ABA will have empty sets of defeasible rules and axiom premises, and consist of ordinary premises and strict rules (respectively corresponding to the assumptions and rules in an ABA theory). Then, for every ordinary premise  $\alpha$ , one specifies that:

- (1) there is a  $\beta$  in  $\mathcal{L}$  such that  $\beta$  is a contrary or contradictory of  $\alpha$
- (2)  $\alpha$  is not the conclusion of a strict inference rule (corresponding to so called ‘flat’ ABA theories)

Then, without the use of preference relation, a correspondence can be shown between ABA and  $ASPIC^+$ . Note that by reconstructing ABA in  $ASPIC^+$ , one can then identify conditions under which ABA satisfies rationality postulates (by requiring, for instance, that the strict rules are closed under transposition). For example, consider the  $ASPIC^+$  reconstruction of an ABA theory consisting of strict rules  $a \rightarrow p$  and  $b \rightarrow \neg p$ , and ordinary premises (assumptions)  $\{a, b\}$  such that  $a$  and  $\neg a$  are contradictories, and  $b$  and  $\neg b$  are contradictories. Consistency is violated since one can construct a single preferred (and grounded) extension containing arguments  $A = [a; a \rightarrow p]$  and  $B = [b; b \rightarrow \neg p]$ , neither of which attack each other. However with the additional transpositions  $p \rightarrow \neg b$  and  $\neg p \rightarrow \neg a$ , then extending  $A$  and  $B$  yields  $A' = [a; a \rightarrow p; p \rightarrow \neg b]$  and  $B' = [b; b \rightarrow \neg p; \neg p \rightarrow \neg a]$ .  $A'$  and  $B'$  respectively attack  $B$  and  $A$ . So the set of arguments  $\{A, B\}$  is no longer admissible (neither  $A$  or  $B$  can defend against these attacks).

The rationale for these more general notions of conflict and attack is two-fold. Firstly, one can for pragmatic reasons state that two formulae are in conflict, rather than requiring that one implies the negation of another; for example, assuming a predicate language with the binary ‘ $<$ ’ relation, one can state that any two formulae of the form  $\alpha < \beta$  and  $\beta < \alpha$  are contradictories. Secondly, the  $\neg$  function allows for an asymmetric notion of negation. This in turn is required for modelling negation as failure (as in logic programming). Using the negation as failure symbol  $\sim$  (also called ‘weak’ negation, in contrast to the ‘strong’ negation symbol  $\neg$ ), then  $\sim \alpha$  denotes the negation of  $\alpha$  under the assumption that  $\alpha$  is not provable (i.e., the negation of  $\alpha$  is assumed in the absence of evidence to the contrary). It is not then meaningful to assert that such an assumption brings into question (and so initiates an attack on) the evidence whose very absence is required to make the assumption in the first place. In other words, if  $A$  is an argument consisting of the premise  $\sim \alpha$ , and  $B$  concludes  $\alpha$  (the contrary of  $\sim \alpha$ ), then  $B$  attacks  $A$ , but not vice versa. Furthermore, since the very construction of  $A$  is invalidated by evidence to the contrary, i.e.,  $B$ , then such attacks succeed as defeats *independently* of preferences.

To accommodate the notion of contrary, and attacks on contraries succeeding as defeats independently of preferences, we further modify Definition 3.10 to distinguish the special cases where  $\text{Conc}(A)$  is a contrary of  $\varphi$ , in which case we say that  $A$  *contrary rebuts*  $B$  and  $A$  *contrary undermines*  $B$ , and then modify Definition 3.12 so that:

- $A$  successfully rebuts  $B$  if  $A$  contrary rebuts  $B$ , or  $A$  rebuts  $B$  on  $B'$  and  $A \not\prec B'$ .
- $A$  successfully undermines  $B$  if  $A$  contrary undermines  $B$ , or  $A$  undermines  $B$  on  $\phi$  and  $A \not\prec \phi$ .

Following on from the discussion in Section 4.2, one can then show (Modgil and

Prakken 2013) that with the additional notion of contrary, satisfaction of the four rationality postulates not only requires that the argument theory satisfy axiom consistency, and transposition or contraposition, but also that it is *well formed* in the sense that whenever  $\phi$  is a contrary of  $\psi$  then  $\psi$  is not an axiom premise or the consequent of a strict rule.

To illustrate the use of negation as failure, suppose you want your arguments to be built from a propositional language that includes both  $\neg$  and  $\sim$ . One could then define  $\mathcal{L}$  as a language of propositional literals, composed from a set of propositional atoms  $\{a, b, c, \dots\}$  and the symbols  $\neg$  and  $\sim$ . Then:

- $\alpha$  is a *strong literal* if  $\alpha$  is a propositional atom or of the form  $\neg\beta$  where  $\beta$  is a propositional atom (strong negation cannot be nested).
- $\alpha$  is a wff of  $\mathcal{L}$ , if  $\alpha$  is a strong literal or of the form  $\sim\beta$  where  $\beta$  is a strong literal (weak negation cannot be nested).

Then  $\alpha \in \overline{\beta}$  iff (1)  $\alpha$  is of the form  $\neg\beta$  or  $\beta$  is of the form  $\neg\alpha$ ; or (2)  $\beta$  is of the form  $\sim\alpha$  (i.e., for any wff  $\alpha$ ,  $\alpha$  and  $\neg\alpha$  are contradictories and  $\alpha$  is a contrary of  $\sim\alpha$ ). Finally, for any  $\sim\alpha$  that is in the antecedent of a strict or defeasible inference rule, one is required to include  $\sim\alpha$  in the ordinary premises.

Consider now Example 3.7, where we now have that  $u \in \overline{\sim u}$ , and we replace the rule  $d_4 : u \Rightarrow v$  with  $d'_4 : \sim u \Rightarrow v$ , and add  $\sim u$  to the ordinary premises:  $\mathcal{K}_p = \{\sim u, s, u, x\}$ . Then, the arguments  $C_3$  and  $D_4$  are now replaced by arguments  $C'_3$  and  $D'_4$  each of which contain the sub-argument  $E : \sim u$  (instead of  $C_1 : u$ ). Then  $C_1 : u$  contrary undermines, and so defeats,  $C'_3$  and  $D'_4$  on  $\sim u$ .

## 6. Conclusion

### 6.1. Relationship to other approaches

We now sketch the relation of  $ASPIC^+$  with the other approaches presented in this special issue.  $ASPIC^+$  is based on two ideas: that conflicts between arguments are sometimes resolved with explicit preferences, and that arguments are built with two kinds of inference rules: strict, or deductive rules, which logically entail their conclusion, and defeasible rules, which only create a presumption in favour of their conclusion. The second idea implies that  $ASPIC^+$  does not primarily see argumentation as inconsistency handling in a given ‘base’ logic: conflicts between arguments may not only arise from the inconsistency of a knowledge base but also from the defeasibility of the reasoning steps in an argument.  $ASPIC^+$  has these features in common with DeLP (Garcia and Simari 2004) but a main difference with DeLP is that the latter is not meant to generate Dung-style abstract argumentation frameworks. Another difference is that DeLP is more computationally oriented and therefore has restricted expressiveness, while  $ASPIC^+$  is foremost meant as a theoretical framework and therefore has very few restrictions on expressiveness.

The two remaining approaches discussed in this issue, namely, assumption-based argumentation (ABA) (Bondarenko et al. 1997, Dung et al. 2009) and classical argumentation (CA) (Besnard and Hunter 2009) can be formalised within the  $ASPIC^+$  framework (where in the latter case,  $ASPIC^+$  additionally accommodates the use of preferences). However, the focus on explicit preferences and defeasible rules is what sets  $ASPIC^+$  apart from these approaches. As stated by Toni in her chapter in this volume, the philosophy behind ABA is to translate preferences and defeasible rules into ABA rules plus ABA assumptions, so that rebutting and undercutting attack and the application of preferences all reduce to premise attack. The idea of this is to keep the formal theory simpler and to make the technical machinery of ABA available for other approaches. We agree that this approach has its merits but note that it is an open question whether  $ASPIC^+$  can in its full generality be translated into ABA. Also, as we noted above, we claim that there is also

some merit in having a theory with explicit notions of rebutting and undercutting attack and preference application, namely, if the aim is to formalise modes of reasoning in a way that corresponds with human modes of reasoning and debate.

Besnard & Hunter also present a simple rule-based argumentation system, which is very similar to an instantiation of *ASPIC*<sup>+</sup> with only propositional literals in its language, only axiom premises, only defeasible inference rules, no undercutters and no preferences. However, this system differs from *ASPIC*<sup>+</sup> in its notions of argument and attack, for which reason the results on how *ASPIC*<sup>+</sup> satisfies the rationality postulates are not inherited by the simple system.

We also note that both Toni's and Besnard & Hunter's use of the terms 'deductive' and 'defeasible' argumentation differs from ours. Toni calls an inference rule defeasible if it is applied to an assumption. Thus she would, for example, call the inference rules of classical logic defeasible when they are applied to assumptions, while we would still call them deductive. Besnard & Hunter define any form of argumentation as deductive in which the premises "entail" the conclusion according to some monotonic "base logic". Since they assume no constraints on what such a base logic is, in fact any form of structured argumentation fits their definition of a base logic, since the idea of argumentation is that the construction of arguments is monotonic and that the nonmonotonicity arises from the interaction of conflicting arguments. In consequence, they would call *ASPIC*<sup>+</sup>'s defeasible arguments deductive since *ASPIC*<sup>+</sup>'s definition of an argument fits their definition of a base logic. By contrast, we call an argument deductive only if it cannot be attacked on its inferences or conclusion. For example, while Besnard & Hunter call their 'simple system' a system for deductive argumentation, we say it is for defeasible argumentation, since it allows inferences to be attacked on their conclusions.

## 6.2. Further reading

*ASPIC*<sup>+</sup> has been further studied and applied in a number of works. Wu (2012) has studied how *ASPIC*<sup>+</sup> can be modified to satisfy several additional rationality postulates proposed by Caminada et al. (2012). van Gijzel and Prakken (2012) have reconstructed the Carneades system of Gordon et al. (2007) in *ASPIC*<sup>+</sup>. Bex et al. (2013) have used *ASPIC*<sup>+</sup> to give a logical account of the Argument Interchange Format (AIF). Moreover, *ASPIC*<sup>+</sup> has been applied to legal reasoning (Prakken 2012, Prakken et al. 2013), to decision making (van der Weide 2011), to policy debates (Bench-Capon et al. 2011, Prakken 2012) and to IT security risk assessment (Prakken et al. 2013). Finally, two implementations are available online of instantiations of *ASPIC*<sup>+</sup> with domain-specific inference rules and with rule priorities: Mark Snaith's TOAST (<http://www.arg.dundee.ac.uk/toast/>) and Wietske Visser's EPR (<http://www.wietskevisser.nl/research/epr/>).

## 6.3. Final remarks

In this paper we have given a tutorial introduction to the *ASPIC*<sup>+</sup> framework for structured argumentation. *ASPIC*<sup>+</sup> is not a system but a framework for specifying systems. A main objective is to identify conditions under which instantiations of *ASPIC*<sup>+</sup> satisfy logical consistency and closure properties. We first discussed *ASPIC*<sup>+</sup>'s philosophical underpinnings. We then illustrated the main definitions with examples and we presented some more and less principled ways to instantiate the framework. We also briefly discussed how *ASPIC*<sup>+</sup> captures several other approaches as special cases. As we saw above, the *ASPIC*<sup>+</sup> framework can be instantiated in many different ways. We have already discussed some of these ways and their properties. We hope that in due course more 'best practices' in using *ASPIC*<sup>+</sup> will emerge.

## References

- Amgoud, L., and Besnard, P. (2009), “Bridging the gap between abstract argumentation systems and logic,” in *Proceedings of the 3rd International Conference on Scalable Uncertainty (SUM’09)*, eds. L. Godo and A. Pugliese, no. 5785 in Springer Lecture Notes in AI, Berlin: Springer Verlag, pp. 12–27.
- Amgoud, L., and Besnard, P. (2013), “Logical limits of abstract argumentation frameworks,” *Journal of Applied Non-classical Logics*, 23, 229–267.
- Amgoud, L., and Cayrol, C. (2002), “A model of reasoning based on the production of acceptable arguments,” *Annals of Mathematics and Artificial Intelligence*, 34, 197–215.
- Bench-Capon, T., Prakken, H., and Visser, W. (2011), “Argument schemes for two-phase democratic deliberation,” in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*, New York: ACM Press, pp. 21–30.
- Besnard, P., and Hunter, A. (2009), “Argumentation based on classical logic,” in *Argumentation in Artificial Intelligence* eds. I. Rahwan and G. Simari, Berlin: Springer, pp. 133–152.
- Bex, F., Modgil, S., Prakken, H., and Reed, C. (2013), “On logical specifications of the Argument Interchange Format,” *Journal of Logic and Computation*, 23, 951–989.
- Bondarenko, A., Dung, P., Kowalski, R., and Toni, F. (1997), “An abstract, argumentation-theoretic approach to default reasoning,” *Artificial Intelligence*, 93, 63–101.
- Brewka, G. (1989), “Preferred Subtheories: An Extended Logical Framework for Default Reasoning,” in *Proceedings of the 11th International Joint Conference on Artificial Intelligence (IJCAI-89)*, pp. 1043–1048.
- Caminada, M., *For the sake of the Argument. Explorations into argument-based reasoning*, Doctoral dissertation Free University Amsterdam (2004).
- Caminada, M., and Amgoud, L. (2007), “On the evaluation of argumentation formalisms,” *Artificial Intelligence*, 171, 286–310.
- Caminada, M., Carnielli, W., and Dunne, P. (2012), “Semi-stable semantics,” *Journal of Logic and Computation*, 22, 1207–1254.
- Dung, P. (1995), “On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and  $n$ -person games,” *Artificial Intelligence*, 77, 321–357.
- Dung, P., Kowalski, R., and Toni, F. (2009), “Assumption-based argumentation,” in *Argumentation in Artificial Intelligence* eds. I. Rahwan and G. Simari, Berlin: Springer, pp. 199–218.
- Garcia, A., and Simari, G. (2004), “Defeasible Logic Programming: An Argumentative Approach,” *Theory and Practice of Logic Programming*, 4, 95–138.
- Gordon, T., Prakken, H., and Walton, D. (2007), “The Carneades model of argument and burden of proof,” *Artificial Intelligence*, 171, 875–896.
- Kraus, S., Lehmann, D., and Magidor, M. (1990), “Nonmonotonic reasoning, preferential models and cumulative logics,” *Artificial Intelligence*, 44, 167–207.
- Modgil, S., and Prakken, H. (2013), “A general account of argumentation with preferences,” *Artificial Intelligence*, 195, 361–397.
- Pearl, J. (1992), “Epsilon-semantics,” in *Encyclopedia of Artificial Intelligence* ed. S. Shapiro, New York: John Wiley & Sons, pp. 468–475.
- Pollock, J., *Cognitive Carpentry. A Blueprint for How to Build a Person*, Cambridge, MA: MIT Press (1995).
- Pollock, J. (2009), “A recursive semantics for defeasible reasoning,” in *Argumentation in Artificial Intelligence* eds. I. Rahwan and G. Simari, Berlin: Springer, pp. 173–197.
- Prakken, H. (2010), “An abstract framework for argumentation with structured arguments,” *Argument and Computation*, 1, 93–124.
- Prakken, H. (2012), “Formalising a legal opinion on a legislative proposal in the ASPIC+ framework,” in *Legal Knowledge and Information Systems. JURIX 2012: The Twenty-fifth Annual Conference* ed. B. Schafer, Amsterdam etc.: IOS Press, pp. 119–128.
- Prakken, H. (2012), “Reconstructing Popov v. Hayashi in a framework for argumentation with structured arguments and Dungean semantics,” *Artificial Intelligence and Law*, 20, 57–82.
- Prakken, H. (2012), “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* Berlin/Heidelberg: Springer, pp. 249–272.
- Prakken, H., Ionita, D., and Wieringa, R. (2013), “Risk assessment as an argumentation game,” in *Proceedings of the 14th International Workshop on Computational Logic in Multi-Agent Systems (CLIMA XIV)*, eds. J. Leite, T. Son, P. Torroni, L. van der Torre and S. Woltran, no. 8143 in Springer Lecture Notes in AI, Berlin: Springer Verlag, pp. 357–373.
- Prakken, H., Wyner, A., Bench-Capon, T., and Atkinson, K. (2013), “A formalisation of argumentation schemes for legal case-based reasoning in ASPIC+,” *Journal of Logic and Computation*, In press.
- van der Weide, T., *Arguing to Motivate Decisions*, Doctoral dissertation Department of Information and Computing Sciences, Utrecht University (2011).
- van Gijzel, B., and Prakken, H. (2012), “Relating Carneades with abstract argumentation via the ASPIC+ framework for structured argumentation,” *Argument and Computation*, 3, 21–47.
- Walton, D., *Argumentation Schemes for Presumptive Reasoning*, Mahwah, NJ: Lawrence Erlbaum Associates (1996).
- Walton, D., *Fundamentals of Critical Argumentation*, Cambridge: Cambridge University Press (2006).
- Wu, Y., *Between Argument and Conclusion. Argument-based Approaches to Discussion, Inference and Uncertainty*, Doctoral Dissertation Faculty of Sciences, Technology and Communication, University of Luxembourg (2012).