1

Abstract Rule-Based Argumentation

Sanjay Modgil, Henry Prakken

ABSTRACT. This chapter reviews abstract rule-based approaches to argumentation, in particular the *ASPIC*$^+$ framework. In *ASPIC*$^+$ and its predecessors, going back to the seminal work of John Pollock, arguments can be formed by combining strict and defeasible inference rules and conflicts between arguments can be resolved in terms of a preference relation on arguments. This results in abstract argumentation frameworks (a set of arguments with a binary relation of defeat), so that arguments can be evaluated with the theory of abstract argumentation. First the basic *ASPIC*$^+$ framework is reviewed, possible ways to instantiate it are discussed and how these instantiations can satisfy closure and consistency properties. Then the relation between *ASPIC*$^+$ and other work in formal argumentation and nonmonotonic logic is discussed, including a review of how other approaches can be reconstructed as instantiations of *ASPIC*$^+$. Further developments and variants of the basic *ASPIC*$^+$ framework are also reviewed, including developments with alternative or generalised notions of attack and defeat and variants with further constraints on arguments. Finally, implementations and applications of *ASPIC*$^+$ are briefly reviewed and some open problems and avenues for further research are discussed.

1 Introduction

One of the oldest research strands in the logical study of argumentation is to allow for arguments that combine strict and defeasible inference rules. Strict inference rules are intended to capture deductively valid inferences, where the truth of the premises guarantees the truth of the conclusion. Defeasible inference rules are instead meant to capture presumptive inferences, where the premises only create a presumption in favour of the conclusion, which can be refuted by evidence to the contrary. This approach was introduced in AI by [Pollock, 1987; Pollock, 1990; Pollock, 1992; Pollock, 1994; Pollock, 1995], previously studied by e.g. [Lin and Shoham, 1989; Simari and Loui, 1992; Vreeswijk, 1997; Prakken and Sartor, 1997] and [Garcia and Simari, 2004] and currently studied by e.g. [Dung and Thang, 2014; Dung, 2014; Dung, 2016] and in work on the *ASPIC*$^+$ framework [Prakken, 2010; Modgil and Prakken, 2013; Modgil and Prakken, 2014; Caminada *et al.*, 2014; Li and Parsons, 2015; Grooters and Prakken, 2016].

While Dung's seminal theory of abstract argumentation frameworks [Dung, 1995] has proved to be extremely influential, it adopts a level of abstraction that precludes provision of guidelines for choosing how to define arguments and attacks from knowledge bases, and a study of how these choices should be made to ensure rational outcomes yielded by evaluation of the justified arguments under Dung's semantics. The above-mentioned work, which partly originates from before Dung's article, addresses these issues. This chapter presents the current consolidation of this research strand: the *ASPIC*$^+$ framework for structured argumentation. The *ASPIC* framework was initially developed as an output of a European Union project on argumentation [Amgoud *et al.*, 2006] and further developed into the *ASPIC*$^+$ framework, initially in [Prakken, 2010], and subsequently in [Modgil and Prakken, 2013]. The principal aims of *ASPIC*$^+$ were to: 1) generalise *ASPIC* so as to provide a natural knowledge representation framework in which to formalise a wide variety of existing and novel instantiations of abstract argumentation frameworks, while; 2) providing guidelines for instantiations that use features typically incorporated at the abstract level of these frameworks; in particular the use of preferences, which were introduced at the abstract level to determine the success of attacks as defeats [Amgoud and Cayrol, 2002], but may violate rationality postulates unless one carefully accounts for their use when instantiating abstract argumentation frameworks.

Importantly, the strict and defeasible inference rules in *ASPIC*$^+$are not part of the logical object language (in which the premises and conclusions of arguments are expressed), but are *metalevel* rules for encoding inference over well-formed formulas in some object level language. Also, the *ASPIC*$^+$ framework abstracts from the nature and origin of the inference rules and from the nature of the language over which they are defined. The resulting abstract nature[1] of *ASPIC*$^+$means that it provides a framework enabling the study of various logical instantiations of abstract argumentation frameworks, and conditions under which the extensions of these frameworks (and hence the defined inference relation over the instantiating knowledge base of logical formulae, identified by the conclusions of justified arguments in extensions) satisfy the rationality postulates in [Caminada and Amgoud, 2007] (for example that the conclusions of arguments in an extension are mutually consistent). In fact, Assumption-Based Argumentation (ABA) ([Bondarenko *et al.*, 1997]), which only has strict rules, can also be regarded as abstract rule-based argumentation, since ABA also abstracts from the nature and origin of its inference rules. However, we will (except for some brief comparisons) not discuss ABA in this chapter, as it is reviewed in another chapter of this handbook. The same holds for a particular instantiation of the rule-based approach: Defeasible Logic Programming.

In a rule-based approach, arguments are formed by chaining applications of inference rules into inference trees or graphs. This approach can be contrasted with approaches defined in terms of logical consequence notions, in which arguments are premises-conclusion pairs where the premises are consistent and imply the conclusion according to the consequence notion of some adopted 'base logic'. Examples of this approach are classical-logic argumen-

[1]The aforementioned features of *ASPIC*$^+$are shared by earlier work in this tradition, such as the work of Pollock and [Vreeswijk, 1997], and justifies the title of this chapter.

tation [Cayrol, 1995; Besnard and Hunter, 2001; Besnard and Hunter, 2008; Gorogiannis and Hunter, 2011] and its generalisation into abstract Tarskian-logic argumentation [Amgoud and Besnard, 2013]. It is important to note that, unlike these logic-based approaches, rule-based approaches in general do not adopt a single base logic but two base logics, one for the strict and one for the defeasible rules. This issue will be discussed in detail in Section 3.1 of this chapter. Moreover, we will review how 'base logic' approaches [Hunter, 2010] can be formalised as instances of $ASPIC^+$ in which the logical language is a full propositional or first-order language and the only inference rules defined over this language are strict, and corresponds to the inference rules of the base logic.

This chapter is organised as follows. In Section 2 we incrementally introduce features of the $ASPIC^+$ framework. We first introduce the basic framework in which arguments are built from strict and or defeasible inference rules, and are grounded in fallible or infallible premises. Various notions of attacks as well as the use of preferences to determine defeats are defined. The basic framework can thus capture rule-based approaches to argumentation of the type dating back to John Pollock's work in formal epistemology, and formalisms for encoding the well-known schemes and critical questions approach to argumentation developed by the informal logic community (notably [Walton, 1996]), and widely used to accommodate more human orientated rather than formal logic based instantiations. We then define a version of $ASPIC^+$ that generalises the standard notion of negation used to identify when the claim of one argument is in conflict with an element in the attacked argument. In this way an asymmetric notion of conflict can be represented that allows for instantiations by logical languages with negation as failure, and the study of formalisms such as ABA as instances of $ASPIC^+$.

In Section 3 we provide guidance on how to choose and define the premises and strict and defeasible rules that comprise $ASPIC^+$ arguments, and the preference relations that are used to determine the success of attacks as defeats. We then specify formal guidelines as to how one should make the aforementioned choices to ensure satisfaction of the rationality postulates in [Caminada and Amgoud, 2007]. We also discuss the extent to which reasoning with defeasible rules and/or preferences can be reduced to reasoning in systems that do not distinguish between strict and defeasible rules, and/or do not use preferences. Finally, we discuss how argument schemes with critical questions can be reconstructed in $ASPIC^+$ as defeasible inference rules.

Section 4 then reviews the relation of $ASPIC^+$ with other works on argumentation and nonmonotonic logic. We show how some existing argumentation formalisms can be reconstructed in the $ASPIC^+$ framework; in particular, ABA as formulated in [Dung *et al.*, 2007], the Carneades system [Gordon *et al.*, 2007; Gordon and Walton, 2009a], and argumentation formalisms based on Tarskian abstract logics [Amgoud and Besnard, 2013] and in particular classical logic argumentation [Gorogiannis and Hunter, 2011]. We will also discuss how the inference relations of existing non-monotonic logics, in particular Preferred Subtheories [Brewka, 1989] and Prioritised Default Logic [Brewka, 1994a], can be endowed with argumentation semantics through instantiation of the *ASPIC*$^+$ framework. We conclude by reviewing how our structured approach to argumentation sheds light on developments of the theory of abstract argumentation frameworks, including the use of preferences and values, support relations, attacks on attacks, resolutions of attacks and the dynamics of abstract argumentation frameworks.

Further developments of the *ASPIC*$^+$ framework will be discussed in Section 5, in particular studies of alternative notions of attack, studies of generalised notions of attack and defeat, and studies of further consistency, minimality and chaining restrictions on arguments. Implementations and applications of *ASPIC*$^+$ are discussed in Section 6 and we conclude with a discussion of open problems and future research directions in Section 7.

2 *ASPIC*$^+$: Defining the Framework

2.1 The underlying ideas

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 on its grounds or its inferences. However, in real life our grounds may not be indisputable and 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. While this reasoning cannot be attacked, the argument is still fallible since 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, challenging our belief that John was in Holland Park that morning, since witnesses usually speak the truth. Maybe we have a supporting reason for our belief that John was in Holland Park; that we went jogging in Holland Park and saw John and that our senses are usually accurate. But given Jan's testimony, perhaps our senses betrayed us? But then we discover 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.

This example is displayed in Figure 1, where the strict inference is visualised with solid lines, the defeasible inferences with dotted lines and the attack relations with arrow. The defeasible inferences within arguments are supposed to

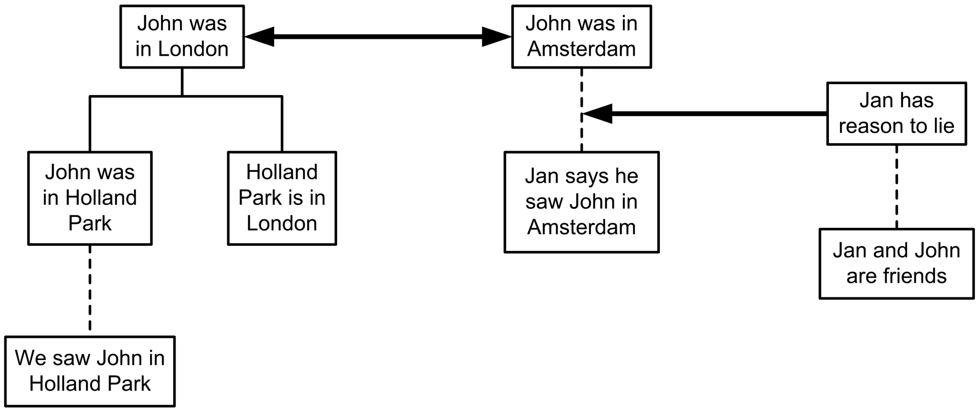

Figure 1. An informal example

be licensed by the generalisations in the example.

If we want to formalise a logic for argumentation, then this simple example already suggests a number of issues to be addressed. 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. $ASPIC^+$ 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) The initial argument was then extended 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 if all 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 subarguments). It is for this very same reason that the third type of attack on a deductive inferential step is also not possible.

$ASPIC^+$ 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 $ASPIC^+$ can be constructed using *defeasible* inference rules, and arguments can be attacked on both the conclusions, and 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.

As well as *fallible* premises that can be attacked, $ASPIC^+$ also allows to distinguish premises that are axiomatic and so cannot be attacked. We discuss the uses of such premises in Section 2.2.1, 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 2.2.1), 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 abstract argumentation 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 here not justify the need for preferences. However, examples are given in the remainder of the paper.

2.2 The basic framework with symmetric negation

2.2.1 Argumentation systems, knowledge bases, and arguments

$ASPIC^+$ is a general framework that allows one to choose a *logical language* $\mathcal{L}$ closed under negation $\neg$ (which we later replace with a more general notion of conflict) and two (possibly empty) sets of *strict* ($\mathcal{R}_s$) and *defeasible* ($\mathcal{R}_d$) inference rules. One also specifies well-formed formulas in $\mathcal{L}$ that correspond to (i.e., name) defeasible rules in $\mathcal{R}_d$ via a partial function n. These names can then be used when attacking arguments on defeasible inference steps. Informally, $n(r)$ is a well-formed formula (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[2].

Definition 2.1 *[Argumentation systems] An* argumentation system *is a triple* $AS = (\mathcal{L}, \mathcal{R}, n)$ *where:*

- $\mathcal{L}$ *is a logical language with a unary negation symbol* $\neg$.
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ *is a set of strict* ($\mathcal{R}_s$) *and defeasible* ($\mathcal{R}_d$) *inference rules of the form* $\varphi_1, \ldots, \varphi_n \rightarrow \varphi$ *and* $\varphi_1, \ldots, \varphi_n \Rightarrow \varphi$ *respectively (where* φ_i, φ *are meta-variables ranging over wff in* $\mathcal{L}$*), and* $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- n *is a partial function such that* $n : \mathcal{R}_d \longrightarrow \mathcal{L}$.

We write $\psi = -\varphi$ *just in case* $\psi = \neg\varphi$ *or* $\varphi = \neg\psi$ *(we will sometimes informally say that formulas* φ *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 $\alpha, \alpha \supset \beta \rightarrow \beta$[3], where α and β 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 3.1.

$ASPIC^+$ also requires that one specify a knowledge base from which the premises of an argument can be taken, where one can distinguish between ordinary premises which are uncertain and so can be attacked, and axiom premises that are certain and so cannot be attacked.

Definition 2.2 *[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*).*

An argumentation theory consists of an argumentation system and a knowledge base:

Definition 2.3 *[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

[2] n is a partial function since you may want to enforce that some defeasible inference steps cannot be attacked.

[3] In this chapter we use $\supset$ to denote the material implication connective of classical logic.

inference graphs, 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 2.4 *[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 any structure obtainable by applying one or more of the following steps finitely many times:*

1. *φ is an argument if $\varphi \in \mathcal{K}$ with:* $\texttt{Prem}(A) = \{\varphi\}$, $\texttt{Conc}(A) = \varphi$, $\texttt{Sub}(A) = \{\varphi\}$, $\texttt{DefRules}(A) = \emptyset$, $\texttt{TopRule}(A) =$ *undefined.*

2. *$A_1, \ldots, A_n \rightarrow \psi$ is an argument if $A_1, \ldots, A_n$ are arguments such that there exists a strict rule*
$\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi$ *in* $\mathcal{R}_s$.
$\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
$\texttt{Conc}(A) = \psi$,
$\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$.
$\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n)$,
$\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \rightarrow \psi$

3. *$A_1, \ldots, A_n \Rightarrow \psi$ is an argument if $A_1, \ldots, A_n$ are arguments such that there exists a defeasible rule* $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi$ *in* $\mathcal{R}_d$.
$\texttt{Prem}(A) = \texttt{Prem}(A_1) \cup \ldots \cup \texttt{Prem}(A_n)$,
$\texttt{Conc}(A) = \psi$,
$\texttt{Sub}(A) = \texttt{Sub}(A_1) \cup \ldots \cup \texttt{Sub}(A_n) \cup \{A\}$,
$\texttt{DefRules}(A) = \texttt{DefRules}(A_1) \cup \ldots \cup \texttt{DefRules}(A_n) \cup \{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi\}$,
$\texttt{TopRule}(A) = \texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n) \Rightarrow \psi$.

Each of these functions `Func` are also defined on sets of arguments $S = \{A_1, \ldots, A_n\}$ as follows: $\texttt{Func}(S) = \texttt{Func}(A_1) \cup \ldots \cup \texttt{Func}(A_n)$. Moreover, for any argument A we define $\texttt{Prem}_n(A) = \texttt{Prem}(A) \cap \mathcal{K}_n$ and $\texttt{Prem}_p(A) = \texttt{Prem}(A) \cap \mathcal{K}_p$.

Example 2.5 *Consider a knowledge base in an argumentation system with $\mathcal{L}$ consisting of $p, q, r, s, t, u, v, x, d_1, d_2, d_3, d_4, d_5$ 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*[4]

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:	$t \Rightarrow \neg d_1$				

[4]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 2.1.

Let $\mathcal{K}_n = \{p\}$ 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.[5]

An argument for r (i.e., with conclusion r) is displayed in Figure 2, 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

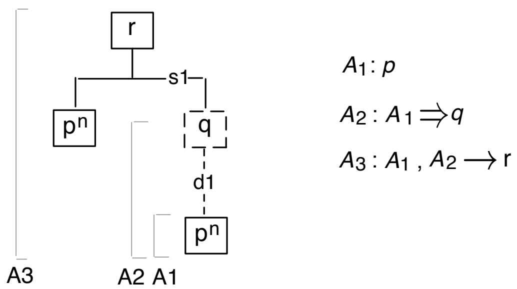

Figure 2. An argument

have that

$\texttt{Prem}(A_3) =$	$\{p\}$	$\texttt{DefRules}(A_3) =$	$\{d_1\}$
$\texttt{Conc}(A_3) =$	r	$\texttt{TopRule}(A_3) =$	s_1
$\texttt{Sub}(A_3) =$	$\{A_1, A_2, A_3\}$		

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

Definition 2.6 *[Argument properties] An argument A is* strict *if* $\texttt{DefRules}(A) = \emptyset$*;* defeasible *if* $\texttt{DefRules}(A) \neq \emptyset$*;* firm *if* $\texttt{Prem}(A) \subseteq \mathcal{K}_n$*;* plausible *if* $\texttt{Prem}(A) \cap \mathcal{K}_p \neq \emptyset$*. An argument is* fallible *if it is defeasible or plausible and* infallible *otherwise. We write $S \vdash \varphi$ if there exists a strict argument for φ with all premises taken from S, and $S \mid\!\sim \varphi$ if there exists a defeasible argument for φ with all premises taken from S.*

[5]In our further examples we will often leave the logical language $\mathcal{L}$ and the n function implicit, trusting that they will be obvious.

Example 2.7 *In Example 2.5 the argument A_1 is both strict and firm, while A_2 and A_3 are defeasible and firm. Furthermore, we have that $\mathcal{K} \vdash p$, $\mathcal{K} \mid\!\sim q$ and $\mathcal{K} \mid\!\sim r$.*

In logic-based approaches to argumentation [Besnard and Hunter, 2008; Amgoud and Besnard, 2013] arguments are often required to be minimal in that no proper subset of their premises should logically (according to the adopted base logic) imply the conclusion. In the *ASPIC*$^+$ context such a constraint would be fine for applications of strict rules and below we will review work that imposes such constraints on *ASPIC*$^+$ arguments (Sections 4.2 and 5.1). However, minimality cannot be required for application of defeasible inference rules, since defeasible rules that are based on more information may well make an argument stronger. For example, *Observations done in ideal circumstances are usually correct* is stronger than *Observations are usually correct.*

Another requirement of logic-based approaches, namely, that an argument's premises have to be consistent, can optionally be imposed in basic *ASPIC*$^+$, leading to two variants of the basic framework. We define a special class of arguments whose premises are 'c-consistent' (for 'contradictory-consistent'). In this way *ASPIC*$^+$ can be used as a framework for reconstructing logic-based argumentation formalisms, as we will further discuss in Section 4.2.

Definition 2.8 *[c-consistency] A set $S \subseteq \mathcal{L}$ is* c-consistent *if for no ϕ is it the case that $S \vdash \phi$ and $S \vdash -\phi$. Otherwise S is said to be* c-inconsistent*. We say that $S \subseteq \mathcal{L}$ is* minimally c-inconsistent *iff S is c-inconsistent and $\forall S' \subset S$, S' is* c-consistent*.*

Definition 2.9 *[c-consistent arguments] An argument A is* c-consistent *iff* $\texttt{Prem}(A)$ *is c-consistent.*

2.2.2 Attack and defeat

ASPIC$^+$ generates abstract argumentation frameworks consisting of arguments related by binary defeats. Having defined arguments above, we now define the attack relation and then apply preferences to determine the defeat relation (in fact [Dung, 1995] called his relation "attack" but we reserve this term for the basic notion of conflict, to which we then apply preferences).

Attack We first present the three ways in which *ASPIC*$^+$ arguments can be in conflict (i.e., attack). 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). In Section 2.1 we argued that arguments cannot be attacked on their strict inferences. In Section 3.3 we will also show that attacks on conclusions of strict inferences may result in violation of rationality postulates. In Section 5.3 we will discuss to what extent alternative definitions of rebutting attack still make sense.

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 $-n(r)$.

Definition 2.10 *[Attacks] A* attacks *B iff A* undercuts, rebuts *or* undermines *B, where:*

- *A* undercuts *argument B (on B') iff $\texttt{Conc}(A) = -n(r)$ for some $B' \in \texttt{Sub}(B)$ such that B''s top rule r is defeasible.*
- *A* rebuts *argument B (on B') iff $\texttt{Conc}(A) = -\varphi$ for some $B' \in \texttt{Sub}(B)$ of the form $B''_1, \ldots, B''_n \Rightarrow \varphi$.*
- *Argument A* undermines *B (on φ) iff $\texttt{Conc}(A) = -\varphi$ for an ordinary premise φ 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 when determining whether attacks succeed as defeats.

Example 2.11 *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:*

B_1: s
B_2: $B_1 \Rightarrow t$
B_3: $B_2 \Rightarrow \neg d_1$

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

C_1: u
C_2: $C_1 \Rightarrow v$
C_3: $C_2 \rightarrow \neg s$

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:

C_1: u
C_2: $C_1 \Rightarrow v$
D_3: x
D_4: $C_2, D_3 \Rightarrow \neg t$

All arguments and attacks in the example are displayed in Figure 3.

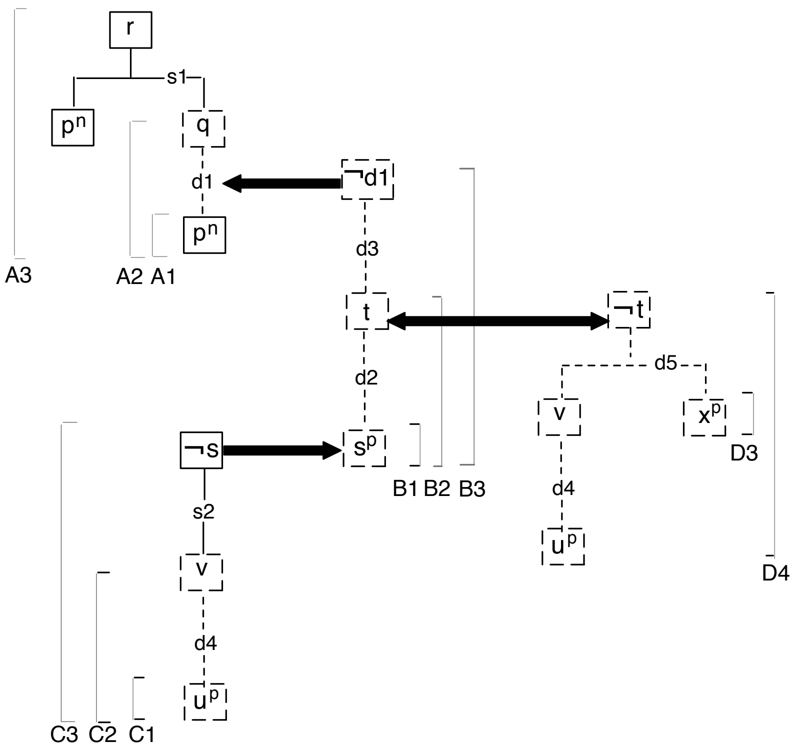

Figure 3. The arguments and attacks in the running example

Defeat The attack relation tells us which arguments are in conflict with each other. If an argument A *successfully attacks*, i.e., defeats, B, then A can be used as a counter-argument to 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. 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). $ASPIC^+$ allows for any strict binary preference ordering $\prec$ on the set of all arguments that can be constructed on the basis of an argumentation theory. Note that in this chapter we formalise argument orderings not as they are defined in [Modgil and Prakken, 2013], but as they are defined in an erratum available online at `https://nms.kcl.ac.uk/sanjay.modgil/AIJfinalErratum`. The erratum essentially reverts to the directly defined strict partial ordering $\prec$ over arguments as employed in [Prakken, 2010]. Then (as illustrated in Section 3.2), the non-strict $\preceq$ is defined so that $A \preceq B$ iff $A \prec B$ or the fallible elements in A and B that are used in deciding preferences, are the same. Moreover, [Modgil and Prakken, 2013] identify conditions under which argument orderings are well-behaved in that they ensure satisfaction of the rationality postulates. The erratum modifies these conditions, which in [Modgil and Prakken, 2013] are stated by reference to non-strict orderings over sets of defeasible rules (ordinary premises), but in the erratum are stated with respect to strict orderings over sets of defeasible rules (ordinary premises). This has been done in order to address a counterexample to rationality pointed out by Sjur Dyrkolbotn (personal communication), assuming the conditions as stated in [Modgil and Prakken, 2013][6]. We will review these conditions later in this chapter.

Definition 2.12 *[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 φ and $A \not\prec \varphi$.*
- A defeats B *iff A undercuts or successfully rebuts or successfully undermines B. (In general, we say A* strictly *defeats B if A defeats B and B does not defeat A).*

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, so the preference $A \not\prec \varphi$ is well defined.

Example 2.13 *In our running example, 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 and D_4 are incomparable, 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 (here we write '*(c-)SAF*' as meaning '*SAF* or *c-SAF*'):

Definition 2.14 *[c-SAFs] Let AT be an argumentation theory (AS, KB). A (c-)structured argumentation framework (* (c-)SAF *) defined by AT, is a triple $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ where*

- *In a SAF, $\mathcal{A}$ is the set of all arguments constructed from KB in AS satisfying Definition 2.4;*

[6]Note that the erratum also addresses a counterexample to rationality in [Dung, 2016].

- *In a c-SAF, $\mathcal{A}$ is the set of all c-consistent arguments constructed from KB in AS satisfying Definition 2.4;*
- *$\preceq$ is a preference ordering on $\mathcal{A}$;*
- *$(X, Y) \in \mathcal{C}$ iff X attacks Y.*

Note that a *c-SAF* is a *SAF* in which all arguments are required to have a c-consistent set of premises.

Example 2.15 *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.*

2.2.3 Generating abstract argumentation frameworks

We now instantiate abstract argumentation frameworks with *ASPIC*$^+$ arguments and defeats.

Definition 2.16 (Argumentation frameworks) *An* abstract argumentation framework (*AF*) corresponding to a *(c-)SAF* $= \langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ *is a pair* $(\mathcal{A}, \mathcal{D})$ *such that $\mathcal{D}$ is the defeat relation on $\mathcal{A}$ determined by $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$.*

The justified arguments of the above defined abstract argumentation frameworks are then defined under various semantics, as in [Dung, 1995]:

Definition 2.17 *[Dung Semantics] Let $(\mathcal{A}, \mathcal{D})$ be an AF and $S \subseteq \mathcal{A}$. Then:*

- *S is* conflict free *iff $\forall X, Y \in S$: $(X, Y) \notin \mathcal{D}$*[7].
- *$X \in \mathcal{A}$ is* acceptable with respect to *S iff $\forall Y \in \mathcal{A}$ such that $(Y, X) \in \mathcal{D}$: $\exists Z \in S$ such that $(Z, Y) \in \mathcal{D}$.*
- *S is an* admissible *set iff S is conflict free and $X \in S$ implies X is acceptable w.r.t. S.*
- *S is a* complete *extension iff S is admissible and if $X \in \mathcal{A}$ is acceptable w.r.t. S then $X \in S$;*
- *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;*

[7]Note that in [Modgil and Prakken, 2013] we motivate the use of the *ASPIC*$^+$ attack relation to define conflict-free sets (a set of arguments is conflict-free if there does not exist an attack between any of its contained arguments), 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 2.17's use of the defeat relation for *both* determining conflict freeness and acceptability of arguments.

- *S is a* stable *extension iff S is conflict free and $\forall Y \notin S$, $\exists X \in S$ s.t. $(X, Y) \in \mathcal{D}$.*

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

It is now also possible to define a consequence notion for well-formed formulas. Several definitions are possible. One is:

Definition 2.18 *[Justified Formulae] A wff $\varphi \in \mathcal{L}$ is* sceptically justified *on the basis of a* (c-)SAF *under semantics T if φ is the conclusion of a sceptically justified argument on the basis of the AF corresponding to the* (c-)SAF *under semantics T, and* credulously justified *on the basis of a* (c-)SAF *under semantics T if φ is not sceptically justified and is the conclusion of a credulously justified argument on the basis of the AF corresponding to the* (c-)SAF *under semantics T.*

An alternative definition of skeptical justification is:

> A wff $\varphi \in \mathcal{L}$ is *sceptically justified* on the basis of the *(c-)SAF* under semantics T if all T-extensions of the AF corresponding to the *(c-)SAF* contain an argument with conclusion φ.

While the original definition of skeptical justification requires that there is one argument for φ that is in all extensions, the alternative definition allows that different extensions contain different arguments for φ. In multiple-extension semantics this can make a difference in, for example, cases with so-called floating conclusions; cf. Example 25 of [Prakken and Vreeswijk, 2002].

Example 2.19 *In our running example, if D_4 strictly defeats B_2, then we have a unique extension in all semantics, namely, $E = \{A_1, A_2, A_3, C_1, C_2, C_3, D_3, D_4\}$. 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 E – but if C_3 does not defeat B_1, then the grounded extension consists of A_1, B_1, C_1, C_2, 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, A_2 and A_3, 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.

2.3 The basic framework with possibly non-symmetric negation

The notion of an argumentation system in Section 2.2.1, assumed a language $\mathcal{L}$ with a unary negation symbol $\neg$, which was used in the definition of conflict-based attack. The standard classical interpretation of $\neg$ licenses a symmetric notion of conflict-based attack, so that an argument consisting of an ordinary premise ϕ or with a defeasible top rule concluding ϕ, *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 [Prakken, 2010; Modgil and Prakken, 2013], accommodates a more general notion of conflict, by defining an argumentation system to additionally include a function $^-$ that, for any wff $\psi \in \mathcal{L}$, specifies the set of wff's that are in conflict with ψ, so that one can define both an asymmetric and symmetric notion of conflict-based attack. Formally:

Definition 2.20 *[$^-$ function] $^-$ is a function from $\mathcal{L}$ to $2^{\mathcal{L}}$, such that:*

- *φ is a* contrary *of ψ if $\varphi \in \overline{\psi}$, $\psi \notin \overline{\varphi}$;*
- *φ is a* contradictory *of ψ (denoted by '$\varphi = -\psi$'), if $\varphi \in \overline{\psi}$, $\psi \in \overline{\varphi}$.*

Now $\texttt{Conc}(A) \in \overline{\varphi}$ ($\texttt{Conc}(A) \in \overline{n(r)}$) replaces $\texttt{Conc}(A) = -\varphi$ ($\texttt{Conc}(A) = -n(r)$) in Definition 2.10's definition of attacks. This induces a generalised notion of an argumentation system as a four-tuple $AS = (\mathcal{L}, ^-, \mathcal{R}, n)$ where $\mathcal{L}$, $\mathcal{R}$ and n are defined as in Definition 2.1 and $^-$ is as just defined. The special case of Definition 2.1 can then be reformulated as the case where $^-$ is defined in terms of classical negation as $\alpha \in \overline{\beta}$ iff α is of the form $\neg\beta$ or β is of the form $\neg\alpha$ (i.e., for any wff α, α and $\neg\alpha$ are contradictories). Below we will continue to refer to the special case with $\neg$ as a triple, leaving the $^-$ function implicit.

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 $^-$ function allows for an asymmetric notion of negation. This enables reconstruction of assumption-based argumentation (ABA) in $ASPIC^+$ (indeed the idea of using a $^-$ function is taken from [Bondarenko *et al.*, 1997]). We briefly review this reconstruction in Section 4.1. Closely related to its use in reconstructing ABA, the contrary function allows for the modelling of 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 α under the assumption that α is not provable (i.e., the negation of α is assumed in the absence of evidence for α). Given this intended reading of $\sim$ it is not 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 α (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 2.10 to distinguish the special cases where $\texttt{Conc}(A)$ is a contrary of φ, in which case we say that A *contrary rebuts* B and A *contrary undermines* B, and then modify Definition 2.12 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 ϕ and $A \not\prec \phi$.

The definition of undercutting attack does not need to be changed.

To illustrate the use of negation as failure, suppose one wants 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:

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

Then $\alpha \in \overline{\beta}$ iff (1) α is of the form $\neg\beta$ or β is of the form $\neg\alpha$; or (2) β is of the form $\sim \alpha$ (i.e., for any wff α, α and $\neg\alpha$ are contradictories and α 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 2.5, 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 $\sim u$.

3 Instantiating the $ASPIC^+$ Framework

$ASPIC^+$ is a framework for specifying systems, and so leaves one fully free to make choices as to the logical language, the strict and defeasible inference rules, the axioms and ordinary premises in a knowledge base, and the argument preference ordering. In this section we discuss various more or less principled ways to make these choices, and then show specific uses of $ASPIC^+$.

3.1 Choosing strict and defeasible rules

3.1.1 Domain specific strict inference rules

$ASPIC^+$ allows the specification of 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.

Example 3.1 *Let $\mathcal{R}_d = \{d_1, d_2\}$ and $\mathcal{R}_s = \{s_1, s_2\}$, where:*

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

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.

In Example 3.1, 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 3.2 (Closure under Transposition) *Let $AT = (AS, \mathcal{K})$ be an argumentation theory. Then AT is* closed under transposition *iff if $\phi_1, \ldots, \phi_n \rightarrow \psi \in \mathcal{R}_s$, then for $i = 1 \ldots n$, $\phi_1, \ldots, \phi_{i-1}, -\psi, \phi_{i+1}, \ldots, \phi_n \rightarrow -\phi_i \in \mathcal{R}_s$.*

In general it is a good idea to ensure that an argumentation theory is closed under transposition, since a strict (deductive) 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 3.3.

3.1.2 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? One then reserves the strict inference rules for general patterns of deductive inference, since one might argue that this is what inference rules are meant for in logic. $ASPIC^+$ therefore allows one to base your strict inference rules (and axioms) on a deductive logic of one's choice. One can do so by choosing a semantics for a particular choice of $\mathcal{L}$ with an associated monotonic notion of semantic consequence, and then letting $\mathcal{R}_s$ be rules that are sound with respect to that semantics. For example, suppose $\mathcal{R}_s$ should conform to classical logic, given a standard propositional (or first-order) language, such that arguments can contain any classically valid inference step over this language. This can be done 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 a propositional language has been chosen, then $\mathcal{R}_s$ can be defined as follows. (where $\vdash_{PL}$ denotes standard propositional-logic consequence). For any finite $S \subseteq \mathcal{L}$ and any $\varphi \in \mathcal{L}$:[8]

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

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, strict rules will be closed under transposition, because of the properties of classical logic.

It should be noted that this way of using a logic as the origin of the strict rule makes some implicit assumptions on the chosen logic, for example that it is compact (everything implied by an infinite set is implied by a finite subset) and satisfies the Cut rule (if S implies φ and $S \cup \{\varphi\}$ implies ψ then S implies ψ). In Section 5.1 we return to this issue.

Let us illustrate the crude approach with a variation of Example 3.1. 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 $\mathcal{R}_s$, including, for example, all inferences instantiating the modus ponens scheme $\varphi, \varphi \supset \psi \rightarrow \psi$. Then the arguments change as follows:

A_1:	$WearsRing$	B_1:	$PartyAnimal$
A_2:	$A_1 \Rightarrow Married$	B_2:	$B_1 \Rightarrow Bachelor$
A_3:	$Married \supset \neg Bachelor$	B_3:	$Married \supset \neg Bachelor$
A_4:	$A_2, A_3 \rightarrow \neg Bachelor$	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 one's choice is to build an existing axiomatic system for the logic into $ASPIC^+$. Its axiom(s) (typically a handful) can be encoded 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,

[8]Although antecedents of rules formally are sequences of formulas, we will sometimes abuse notation and write them as sets.

modus ponens (i.e, $\varphi \supset \psi, \varphi \rightarrow \psi$)[9]. 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 φ 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 $Bachelor$.

Note that in the sophisticated method, closure under transposition may not hold; our example above does not contain modus tollens (that is, $\varphi \supset \psi, -\psi \rightarrow -\varphi$). However, this desirable form of reasoning can also be enforced without explicitly transposing rules. Recall that $S \vdash \varphi$ was defined as 'there exists a strict argument for φ 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 (if $\vdash$ corresponds to classical provability, as enforced by our choice of axioms and inference rules, then $\vdash$ does indeed contrapose).

Definition 3.3 *[Closure under Contraposition] 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 ϕ, if $S \vdash \phi$, then $S \backslash \{s\} \cup \{-\phi\} \vdash -s$.*

Again, as will be discussed in Section 3.3, closure under contraposition also ensures satisfaction of rationality postulates.

3.1.3 Choosing defeasible inference rules

Regarding the choice of defeasible rules, the question as to whether these can be derived from a logic of our choice, just as with strict rules, is controversial. Some philosophers argue that all rule-like structures that we use in daily life are "inference licenses" and so cannot be expressed in the logical object language. In this view, all defeasible generalisations are inference rules, whether they are domain-specific or not, and are applied to formulas from $\mathcal{L}$ to support new formulas from $\mathcal{L}$.

Others (usually logicians) take a more standard-logic approach (e.g. [Kraus *et al.*, 1990; Pearl, 1992]) whereby all contingent knowledge should be expressed in the object language, and so they reject the idea of domain-specific defeasible inference rules (for the same reason they don't like domain-specific strict rules). They introduce a new connective, e.g., $\rightsquigarrow$, into $\mathcal{L}$ where (informally) $p \rightsquigarrow q$ is read as "If p then normally/typically/usually q". They then want to give a model-theoretic semantics for this connective just as logicians give a

[9]As explained above, this strictly speaking is not a rule but a scheme, with meta variables ranging over $\mathcal{L}$.

model-theoretic semantics for all connectives, except that semantics for these defeasible conditionals focus on a *preferred class* of models (e.g., all models where things are as normal as possible) instead of *all* models of a theory as in semantics for deductive logics. Hence, the model-theoretic interpretation of $p \supset q$ is that q is true in *all* models of p, whereas the model theoretic interpretation of $p \rightsquigarrow q$ is that q is true in all *preferred* models of p.

What inference rules for $\rightsquigarrow$ could result from such an approach? On two things there is consensus: 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 $\varphi \rightsquigarrow \psi \rightarrow -\psi \rightsquigarrow -\varphi$ 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[10].

Let us illustrate the difference between the two approaches, by including defeasible modus ponens for $\rightsquigarrow$ in $\mathcal{R}_d$, and replacing the defeasible inference rules d_1 and d_2 (in Example 3.1) with object-level conditionals expressed in $\mathcal{L}$ and included in $\mathcal{K}_p$:

$$\begin{array}{l} \mathit{WearsRing} \rightsquigarrow \mathit{Married} \in \mathcal{K}_p \text{ and } \mathit{PartyAnimal} \rightsquigarrow \mathit{Bachelor} \in \mathcal{K}_p \\ \mathcal{R}_d = \{\varphi \rightsquigarrow \psi, \varphi \Rightarrow \psi\} \end{array}$$

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

$$\begin{array}{llll} A_1: & \mathit{WearsRing} & B_1: & \mathit{PartyAnimal} \\ A_2: & \mathit{WearsRing} \rightsquigarrow \mathit{Married} & B_2: & \mathit{PartyAnimal} \rightsquigarrow \mathit{Bachelor} \\ A_3: & A_1, A_2 \Rightarrow \mathit{Married} & B_3: & B_1, B_2 \Rightarrow \mathit{Bachelor} \\ A_4: & \mathit{Married} \supset \neg \mathit{Bachelor} & B_4: & \mathit{Married} \supset \neg \mathit{Bachelor} \\ A_5: & A_3, A_4 \rightarrow \neg \mathit{Bachelor} & B_5: & B_3, B_4 \rightarrow \neg \mathit{Married} \end{array}$$

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

Concluding, if desired, at least some of the choices concerning defeasible inference rules can be based 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. 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 3.5.

[10]See Chapter 4 of [Caminada, 2004] for a very readable overview of the discussion.

3.2 Choosing argument preference orderings

A well studied use of preferences in the non-monotonic logic literature is based on the use of preference orderings over formulae in the language or defeasible inference rules. If $ASPIC^+$ is to be used as a framework for giving argumentation-based characterisations of non-monotonic formalisms augmented with preferences, then it needs to provide an account of how these preference orderings can be 'lifted' to preferences over arguments. Since $ASPIC^+$ uses defeasible inference rules and ordinary premises, both may come equipped with preference orderings $\leq$ on $\mathcal{R}_d$ and $\leq'$ on $\mathcal{K}_p$, which in general may be distinct, in keeping with the ontologically distinct nature of rules and premises. For example, the ordinary premises may represent the content of percepts from sensors or of witness testimonies, whose preference ordering reflects the relative reliability of the sensors, respectively witnesses. The defeasible rules may, for example, be ordered 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 preferences 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 partial preorders $\leq$ on $\mathcal{R}_d$ and $\leq'$ on $\mathcal{K}_p$, where as usual, $X <^{(')} Y$ iff $X \leq^{(')} Y$ and $Y \not\leq^{(')} X$ (note that we may represent these orderings in terms of the strict counterpart they define). However, these preferences 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 strict set ordering $\lhd_{\mathtt{s}}$ over sets of rules/premises, where for any sets of defeasible rules/ordinary premises S and S', we intuitively want that:

1) if S is the empty set, it cannot be that $S \lhd_{\mathtt{s}} S'$;

2) if S' is the empty set, it must be that $S \lhd_{\mathtt{s}} S'$ 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$\lhd_{\mathtt{s}}$; the so called `Elitist` and `Democratic` ways (i.e., $\mathtt{s} = \mathtt{Eli}$ and `Dem` respectively). `Eli` compares sets on their minimal and `Dem` on their maximal elements.

Definition 3.4 *[Orderings $\lhd_{\mathtt{s}}$] Let Γ and Γ' be finite sets*[11]*. Then $\lhd_{\mathtt{s}}$ is defined as follows:*

[11] Notice that it suffices to restrict $\lhd$ to finite sets since $ASPIC^+$ arguments are assumed to be finite (in Definition 2.14) and so their sets of ordinary premises/defeasible rules must be finite.

1. *If* $\Gamma = \emptyset$ *then* $\Gamma \not\triangleleft_{\mathtt{s}} \Gamma'$ *;*

2. *If* $\Gamma' = \emptyset$ *and* $\Gamma \neq \emptyset$ *then* $\Gamma \triangleleft_{\mathtt{s}} \Gamma'$ *;*
   else, assuming a preordering $\leq$ *over the elements in* $\Gamma \cup \Gamma'$*, then if :*

3. $\mathtt{s} = \mathtt{Eli}$*:*
   $\Gamma \triangleleft_{\mathtt{Eli}} \Gamma'$ *if* $\exists X \in \Gamma$ *s.t.* $\forall Y \in \Gamma'$, $X < Y$.
   else, if:

4. $\mathtt{s} = \mathtt{Dem}$*:*
   $\Gamma \triangleleft_{\mathtt{Dem}} \Gamma'$ *if* $\forall X \in \Gamma$, $\exists Y \in \Gamma'$, $X < Y$.

For $\mathtt{s} = \mathtt{Eli}$ *or* $\mathtt{s} = \mathtt{Dem}$*:* $\Gamma \trianglelefteq_{\mathtt{s}} \Gamma'$ *iff* $\Gamma = \Gamma'$ *or* $\Gamma \triangleleft_{\mathtt{s}} \Gamma'$

Now the **last-link principle** strictly prefers an argument A over another argument B if the last defeasible rules used in B are less preferred ($\triangleleft_{\mathtt{s}}$) 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 3.5 *[Last defeasible rules] Let A be an argument.*

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

For example, letting $\mathcal{K} = \{p, q\}$, $\mathcal{R}_s = \{r, s \to t\}$ and $\mathcal{R}_d = \{p \Rightarrow r;\ q \Rightarrow s\}$, then
$\mathtt{LastDefRules}(A) = \{p \Rightarrow r;\ q \Rightarrow s\}$ where A is the argument for t.

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

Definition 3.6 *[Last link principle] Let A and B be two arguments. Then $A \prec B$ iff:*

1. $\mathtt{LastDefRules}(A) \triangleleft_{\mathtt{s}} \mathtt{LastDefRules}(B)$*; or*

2. $\mathtt{LastDefRules}(A)$ *and* $\mathtt{LastDefRules}(B)$ *are empty and* $\mathtt{Prem_p}(A) \triangleleft_{\mathtt{s}} \mathtt{Prem_p}(B)$.

Then $B \preceq A$ *iff* $B \prec A$ *or, if* $\mathtt{LastDefRules}(A) \neq \emptyset$ *then* $\mathtt{LastDefRules}(A) = \mathtt{LastDefRules}(B)$, *else* $\mathtt{Prem_p}(A) = \mathtt{Prem_p}(B)$.

Example 3.7 *Suppose in our running example that $u <' s$, $x <' s$, $d_2 < d_5$ and $d_4 < d_2$. Applying the last-link ordering to check whether C_3 defeats B_1, we compare* $\mathtt{LastDefRules}(C_3) = \{d_4\}$ *with* $\mathtt{LastDefRules}(B_1) = \emptyset$. *Clearly, $\{d_4\} \triangleleft_{\mathtt{Eli}} \emptyset$, so $C_3 \prec B_1$, so C_3 does not defeat B_1. Next, to check whether D_4*

defeats B_2, we compare $\texttt{LastDefRules}(B_2) = \{d_2\}$ *with* $\texttt{LastDefRules}(D_4) = \{d_5\}$. *Since $d_2 < d_5$ we have that* $\texttt{LastDefRules}(B_2) \triangleleft_{\texttt{Eli}} \texttt{LastDefRules}(D_4)$, *so D_4 strictly defeats B_2.*

The **weakest-link principle** considers not the last but *all* uncertain elements in an argument.

Definition 3.8 *[Weakest link principle] Let A and B be two arguments. Then $A \prec B$ iff*

1. *If both B and A are strict, then* $\texttt{Prem}_{\texttt{p}}(A) \triangleleft_{\texttt{s}} \texttt{Prem}_{\texttt{p}}(B)$, *else;*
2. *If both B and A are firm, then* $\texttt{DefRules}(A) \triangleleft_{\texttt{s}} \texttt{DefRules}(B)$, *else;*
3. $\texttt{Prem}_{\texttt{p}}(A) \triangleleft_{\texttt{s}} \texttt{Prem}_{\texttt{p}}(B)$ *and* $\texttt{DefRules}(A) \triangleleft_{\texttt{s}} \texttt{DefRules}(B)$

Then $B \preceq A$ iff $B \prec A$ or, $\texttt{DefRules}(A) = \texttt{DefRules}(B)$ *and* $\texttt{Prem}_{\texttt{p}}(A) = \texttt{Prem}_{\texttt{p}}(B)$.

Example 3.9 *In our running example to check whether C_3 defeats B_1 according to the weakest-link ordering, we first compare* $\texttt{Prem}_p(C_3) = \{u\}$ *with* $\texttt{Prem}_p(B_1) = \{s\}$. *Since $u <' s$ we have that* $\texttt{Prem}_p(C_3) \triangleleft_{\texttt{Eli}} \texttt{Prem}_p(B_1)$. *Also,* $\texttt{DefRules}(C_3) = \{d_4\} \triangleleft_{\texttt{Eli}} \texttt{DefRules}(B_1) = \emptyset$, *and so $C_3 \prec B_1$ and C_3 does not defeat B_1.*
For B_2 and D_4: $\texttt{Prem}_p(D_4) = \{u, x\} \triangleleft_{\texttt{Eli}} \texttt{Prem}_p(B_2) = \{s\}$ *since $u <' s$ and $x < s'$. Then since $d_4 < d_2$,* $\texttt{DefRules}(D_4) = \{d_4, d_5\} \triangleleft_{\texttt{Eli}} \texttt{DefRules}(B_2)\{d_2\}$. *So $D_4 \prec B_2$ and B_2 strictly defeats D_4.*

We next present two examples illustrating the suitability of the last-, respectively, weakest-link orderings. Consider an example relating to whether people misbehaving in a university library may be denied access to the library.[12]

Example 3.10 *Let $\mathcal{K}_p = \{Snores,\ Professor\}$, $\mathcal{R}_d =$*

$\{Snores \Rightarrow_{d_1} Misbehaves;$
$Misbehaves \Rightarrow_{d_2} AccessDenied;$
$Professor \Rightarrow_{d_3} \neg AccessDenied\}.$

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

A_1:	*Snores*	B_1:	*Professor*
A_2:	$A_1 \Rightarrow$ *Misbehaves*	B_2:	$B_1 \Rightarrow \neg$*AccessDenied*
A_3:	$A_2 \Rightarrow$ *AccessDenied*		

[12] In all examples below, sets that are not specified are assumed to be empty. Moreover, sometimes we will attach the rule names to the $\Rightarrow$ symbol. Note that these attached indices have no formal meaning and are for ease of reference only.

Let us apply the ordering to the arguments A_3 and B_2. The rule sets to be compared are $\texttt{LastDefRules}(A_3) = \{d_2\}$ *and* $\texttt{LastDefRules}(B_2) = \{d_3\}$. *Since* $d_3 < d_2$ *we have that* $\texttt{LastDefRules}(B_2) \lhd_{\texttt{Eli}} \texttt{LastDefRules}(A_3)$, *hence* $B_2 \prec A_3$. *So A_3 strictly defeats B_2, hence A_3 is justified in any semantics, and 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 3.8. Since Snores $<'$ Professor, we have that $\texttt{Prem}_p(A_3) \lhd_{\texttt{Eli}} \texttt{Prem}_p(B_2)$. *Furthermore, the rule sets to be compared are now* $\texttt{DefRules}(A_3) = \{d_1, d_2\}$ *and* $\texttt{DefRules}(B_2) = \{d_3\}$. *Since* $d_1 < d_3$ *we have that* $\texttt{DefRules}(A_3) \lhd_{\texttt{Eli}} \texttt{DefRules}(B_2)$. *So now we have that* $A_3 \prec B_2$. *Hence B_2 now strictly defeats A_3 and we conclude instead that ¬AccessDenied.*

Which outcome 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, in an example of exactly the same form, with the legal rules replaced by empirical generalisations, intuitions seem to favour the weakest-link ordering:

Example 3.11 *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 ¬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.

3.3 The rationality postulates of Caminada and Amgoud (2007) and their satisfaction in $ASPIC^+$

$ASPIC^+$ leaves one fully free to choose a language, what is an axiom and what is an ordinary premise and how to specify 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 in the sense that the desirable properties proposed by [Caminada and Amgoud, 2007] are satisfied. Before presenting these properties, we define required notions of direct and indirect consistency in terms of the contrary function (recall Definition 2.20).

Definition 3.12 *[Direct and Indirect Consistency] 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 $\psi \in \overline{\varphi}$*
- indirectly consistent *iff $Cl(S)$ is directly consistent.*

Let E be any complete extension of an abstract argumentation framework corresponding to a (c)-SAF as defined in Section 2.2.3.

> **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 $A' \in \texttt{Sub}(A)$ then $A' \in E$.
>
> **Closure under Strict Rules**: If E contains arguments with conclusions $\alpha_1, \ldots .\alpha_n$, then any arguments obtained by applying only strict inference rules to these conclusions, are in E, i.e., $\{\texttt{Conc}(A) | A \in E\} = Cl(\{\texttt{Conc}(A) | A \in E\})$.
>
> **Direct Consistency**: The conclusions of arguments in E are directly consistent, i.e., $\{\texttt{Conc}(A) | A \in E\}$ is consistent.
>
> **Indirect Consistency**: The conclusions of arguments in E are indirectly consistent, i.e., $Cl(\{\texttt{Conc}(A) | A \in E\})$ is consistent.

We next review the work done on identifying sufficient conditions for $ASPIC^+$ satisfying [Caminada and Amgoud, 2007]'s four rationality postulates.

3.3.1 The work of Caminada and Amgoud (2007), Prakken (2010) and Modgil and Prakken (2013)

The first relevant condition is that an argumentation theory is closed under transposition or contraposition. If neither is satisfied, then since strict rule applications cannot be attacked, direct consistency may be violated. Consider our first version of Example 3.1. Suppose we only have the strict rule s_1 so that B_3 cannot be constructed (given the absence of 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 (under all semantics) which contains arguments for both *Bachelor* and $\neg$*Bachelor* and so violates direct and indirect consistency. However, with transposition we also have s_2. Then B_3 can be constructed, which rebuts A_3 on A_2. Under the last-link ordering (assuming again that $d_1 < d_2$) we still have that A_3 does not defeat B_2, but now B_3 strictly defeats A_2. We have a unique extension in all semantics, containing all arguments except A_2 and A_3. This extension does not violate consistency.

One might argue that the above violation of consistency, before inclusion of the transposed rule s_2, arises because $ASPIC^+$ forbids attacks on strictly derived conclusions. Consistency would not be violated if B_2 was allowed to attack A_3. However, apart from the reasons discussed in Section 2.2.2, another reason for prohibiting attacks on strictly derived conclusions is that if allowed, extensions may not be strictly closed or indirectly consistent, even if the strict rules are closed under transposition. To see why, suppose we 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 3.1. Suppose also that all knowledge-base items and defeasible rules are of equal preference, and 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 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 expressed in the following definition of a 'well-defined' structured argumentation framework (recall Definition 2.14), which references the notion of a 'reasonable' preference relation that is subsequently explained and defined:

Definition 3.13 *[Well defined* (c-)SAF*s] A (c-)SAF* $(\mathcal{A}, \mathcal{C}, \preceq)$ *defined by an an argumentation theory* $AT = (AS, \mathcal{K})$*, where* $AS = (\mathcal{L}, ^-, \mathcal{R}, n)$ *and* $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p$*, is said to be* well defined *iff:*

- AT *is* closed under transposition *or* closed under contraposition.
- $Cl_{\mathcal{R}_s}(\mathcal{K}_n)$ *is consistent (in which case* $\mathcal{K}$ *is said to be* axiom consistent*).*
- *If* $\mathcal{A}$ *is restricted to be the set of c-consistent arguments, then* $\mathcal{A}$ *is* c-classical. *That is to say, for any minimal c-inconsistent* $S \subseteq \mathcal{L}$ *and for any* $\varphi \in S$*, it holds that* $S \setminus \{\varphi\} \vdash -\varphi$ *(i.e., amongst all arguments defined there exists a strict argument with conclusion* $-\varphi$ *with all premises taken from* $S \setminus \{\varphi\}$*).*
- well formed *if whenever* φ *is a contrary of* ψ *then:*
  - $\psi \notin \mathcal{K}_n$*; and*
  - ψ *is not the consequent of a strict rule.*
- $\preceq$ *is reasonable.*

The property of transposition (and the alternative contraposition) has been discussed above. That the axiom premises are required to be consistent when closed under strict rules is self-evident given that axiom premises represent indisputable information or axioms of a deductive logic. The c-classicality condition is only required to hold when using $ASPIC^+$ to reconstruct Tarskian logic, and in particular classical logic approaches to argumentation, where $\mathcal{A}$ is restricted to arguments with consistent premises. Intuitively, c-classicality says that for every minimally c-inconsistent set of wff and any of its elements the remaining maximally c-consistent subset gives rise to an argument against the element. The intuition underlying the well-formed property should be apparent given the motivation for use of the contrary function and preference independent attacks on contraries, as discussed in Section 2.3. We now elaborate on the notion of reasonable preference orderings.

Before doing so, we define the following notion of strict continuations of arguments, which we define in a slightly different way than [Modgil and Prakken, 2013]. The new definition is arguably simpler but does not affect the proofs of Modgil and Prakken. It identifies arguments that are formed by extending a set of arguments with only strict inferences into a new argument, so that the new argument can only be attacked on the arguments that it extends.

Definition 3.14 *[Strict continuations] The set of* strict continuations *of a set of arguments from $\mathcal{A}$ is the smallest set satisfying the following conditions:*

1. *Any argument A is a strict continuation of $\{A\}$.*
2. *If $A_1, \ldots, A_n$ and $S_1, \ldots, S_n$ are such that for each $i \in \{1, \ldots, n\}$, A_i is a strict continuation of S_i and $\{B_{n+1}, \ldots, B_m\}$ is a (possibly empty) set of strict-and-firm arguments, and $\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n), \texttt{Conc}(B_{n+1}), \ldots, \texttt{Conc}(B_m) \rightarrow \varphi$ is a strict rule in R_s, then $A_1, \ldots, A_n, B_{n+1}, \ldots, B_m \rightarrow \varphi$ is a strict continuation of $S_1 \cup \ldots \cup S_n$.*

If argument A is a strict continuation of arguments $\{A_1, \ldots, A_n\}$, then A is a strict argument over $\{\texttt{Conc}(A_1), \ldots, \texttt{Conc}(A_n)\}$.

Example 3.15 In Example 2.5 (see Figure 3) all arguments are strict continuations of the singleton set containing themselves while A_3 is a strict continuation of $\{A_1, A_2\}$ and C_3 is a strict continuation of $\{C_2\}$.

Definition 3.16 *[Reasonable Argument Orderings] An argument ordering $\preceq$ is* reasonable *iff:*

1. *i) $\forall A, B$, if A is strict and firm and B is plausible or defeasible, then $B \prec A$;*
   ii) $\forall A, B$, if B is strict and firm then $B \not\prec A$;
   iii) $\forall A, A', B$ such that A' is a strict continuation of $\{A\}$, if $A \not\prec B$ then $A' \not\prec B$, and if $B \not\prec A$ then $B \not\prec A'$ (i.e., applying strict rules to a single argument's conclusion and possibly adding new axiom premises does not weaken, respectively strengthen, arguments).

2. *Let $\{C_1, \ldots, C_n\}$ be a finite subset of $\mathcal{A}$, and for $i = 1 \ldots n$, let $C^{+\backslash i}$ be some strict continuation of $\{C_1, \ldots, C_{i-1}, C_{i+1}, \ldots, C_n\}$. Then it is not the case that: $\forall i$, $C^{+\backslash i} \prec C_i$.*

A reasonable argument ordering essentially amounts to requiring that: arguments that are both strict and firm are strictly preferred over all plausible or defeasible arguments, and no argument is strictly preferred to a strict and firm argument (1i) and 1ii)); the strength (and implied relative preference) of an argument is determined exclusively by the defeasible rules and/or ordinary premises (1iii)); the preference ordering is acyclic (2).

Indeed, a strict relation $\lhd$ (on sets of ordinary premises or defeasible rules) results in a preference ordering (under either weakest- or last-link) that is reasonable, if $\lhd$ satisfies the following conditions:

Definition 3.17 *[Inducing reasonable orderings] $\lhd$ is said to be* reasonable inducing *if $\lhd$ is a strict partial ordering (irreflexive and transitive) such that:*

for any $\mathtt{kr} \in \{\mathtt{LastDefRules}, \mathtt{DefRules}, \mathtt{Prem_p}\}$, for all arguments $B_1, \ldots, B_n, A$
such that $\bigcup_{i=1}^{n} \mathtt{kr}(B_i) \lhd \mathtt{kr}(A)$, it holds that for some $i = 1 \ldots n$, $\mathtt{kr}(B_i) \lhd \mathtt{kr}(A)$

It can be shown that both $\lhd_{\mathtt{Eli}}$ and $\lhd_{\mathtt{Dem}}$ (recall Definition 3.4) are reasonable inducing.

We are now in a position to state some important results proved in [Modgil and Prakken, 2013]. Any (c)-structured argumentation framework satisfies the rationality postulate of sub-argument closure. Moreover, if a (c-)structured argumentation framework is well-defined then the postulates of strict closure and direct and indirect consistency are also satisfied by the *ASPIC*$^+$ framework as defined with the contrary function in Section 2.3.

Theorem 3.18 *[Sub-argument Closure] Let $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ be a* (c-)SAF *and E a complete extension of the AF corresponding to Δ. Then for all $A \in E$: if $A' \in \mathtt{Sub}(A)$ then $A' \in E$.*

Theorem 3.19 *Let $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$ be a well-formed* (c-)SAF *and E a complete extension of the AF corresponding to Δ. Then*

Closure under Strict Rules $\{\mathtt{Conc}(A) | A \in E\} = Cl_{R_s}(\{\mathtt{Conc}(A) | A \in E\})$;

Direct consistency $\{\mathtt{Conc}(A) | A \in E\}$ *is consistent;*

Indirect consistency $Cl_{R_s}(\{\mathtt{Conc}(A) | A \in E\})$ *is consistent.*

Finally, note that if no strict rules or axiom premises are included in the argumentation theory, then the preference ordering need *not* be reasonable in order for all four rationality postulates to be satisfied (indeed no assumptions as to the properties of the preference ordering are required in this case). Thus the requirement that the defined (c-)SAF be well-defined does not apply.

3.3.2 The work of Dung and Thang (2014) and Grooters (2014)

For the case without preferences and knowledge bases, [Dung and Thang, 2014] identify weaker conditions for satisfying the rationality postulates than those discussed above. [Dung and Thang, 2014] formulate their results in terms of an adaptation of [Amgoud and Besnard, 2013] abstract-logic approach to abstract argumentation with abstract attack and support relations between arguments. After defining their adaptation they apply it to what they call "rule-based systems", which are a pair of sets of strict and defeasible rules defined over a propositional literal language. Since they adopt the $ASPIC^+$ definitions of an argument and of defeat (which they call 'attack') they thus effectively study a class of $ASPIC^+$ instantiations with an empty knowledge base and with no preferences. Below we summarise their definitions and results as holding for this class of $ASPIC^+$ instantiations, adapting fragments of [Grooters, 2014] and [Grooters and Prakken, 2016]. In doing so, we implicitly assume a given $ASPIC^+$ structured argumentation framework generated by a rule-based instantiation in the sense of [Dung and Thang, 2014], which we will call a 'rule-based' $ASPIC^+$ SAF.

First, an argument A is a *basic defeasible argument* iff $\texttt{TopRule}(A) \in \mathcal{R}_d$, and a set X of arguments is called *inconsistent* if $\texttt{Conc}(X)$ is indirectly inconsistent.

Definition 3.20 *[Base of an argument] Let A be an argument and BA a finite set of subarguments of A. BA is a* base *of A if*

- $\texttt{Conc}(A) \in Cl_{\mathcal{R}_s}(\texttt{Conc}(BA))$*;*
- *For each argument C, C defeats A if and only if C defeats BA.*

The following example shows the intuitive idea of a base.

Example 3.21 Let $\mathcal{R}_s = \{c \rightarrow d\}$ and $\mathcal{R}_d = \{\Rightarrow a; \Rightarrow b; a, b \Rightarrow c\}$. Then the following arguments can be constructed: $A_1 :\Rightarrow a$, $A_2 :\Rightarrow b$, $A_3 : A_1, A_2 \Rightarrow c$ and $A_4 : A_3 \rightarrow d$. See Figure 4.

A_4 can only be attacked on its subarguments A_1, A_2, or A_3 because of the strict top rule. Every argument that attacks A_1 or A_2 also attacks A_3, so every argument that attacks A_4 also attacks A_3. It is easy to see that every argument that attacks A_3 also attacks A_4. $\texttt{Conc}(A_4) \subseteq Cl_{\mathcal{R}_s}(\texttt{Conc}(A_3))$, so $\{A_3\}$ is a base of A_4. The same kind of reasoning applies to the fact that the set $\{A_1, A_2, A_3\}$ is also a base of A_4.
However note that the set $\{A_1, A_2\}$ is not a base of A_4, because A_4 can be rebutted (on A_3) without A_1 or A_2 being attacked.

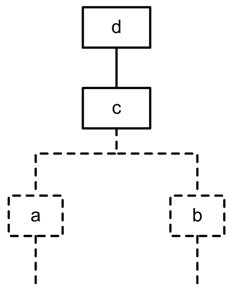

Figure 4. Arguments of Example 3.21

Definition 3.22 *[Generation of arguments] An argument A is said to be* generated *by a set of arguments S, if there is a base B of A such that $B \subseteq \texttt{Sub}(S)$. The set of all arguments generated by S is denoted by $GN(S)$.*

[Dung and Thang, 2014] show that for every set of arguments S, $\texttt{Sub}(S) \subseteq GN(S)$ and for every complete extension E, $GN(E) = E$. [Grooters, 2014] notes that these results immediately imply that each rule-based *ASPIC*$^+$ *SAF* satisfies the closure under subarguments postulate, since for every complete extension E: $\texttt{Sub}(E) \subseteq GN(E) = E$ ([Dung and Thang, 2014] do not consider the subargument-closure postulate).

Definition 3.23 *[Compact] A rule-based* ASPIC$^+$ *SAF is* compact *if for each set of arguments S, $GN(S)$ is closed under strict rules.*

[Dung and Thang, 2014] show that each rule-based *ASPIC*$^+$ *SAF* is compact and that each compact rule-based *ASPIC*$^+$ *SAF* satisfies strict closure, so each rule-based *ASPIC*$^+$ *SAF* satisfies the closure under strict rules postulate.

Definition 3.24 *[Cohesive] A rule-based* ASPIC$^+$ *SAF is* cohesive *if for each inconsistent set of arguments S, $GN(S)$ is conflicting (attacks itself).*

Definition 3.25 *[Self-contradiction axiom] A rule-based* ASPIC$^+$ *SAF is said to satisfy the* self-contradiction axiom *if for each minimal inconsistent set $X \subseteq \mathcal{L}$: $\neg X \subseteq Cl_{\mathcal{R}_s}(X)$ (where $\neg X = \{\neg l \mid l \in L\}$).*

[Dung and Thang, 2014] then show that each cohesive rule-based *ASPIC*$^+$ *SAF* satisfies the indirect-consistency postulate and, moreover, that each rule-based *ASPIC*$^+$ *SAF* that satisfies the self-contradiction axiom is cohesive. Combining these two results, it follows that each rule-based *ASPIC*$^+$ *SAF* that satisfies the self-contradiction axiom, also satisfies indirect consistency. This result generalises the corresponding results discussed in the previous subsection, since satisfying the self-contradiction axiom is a weaker notion than closure under transposition. First, [Dung and Thang, 2014] prove that the latter implies the former in that each rule-based *ASPIC*$^+$ *SAF* that is closed under transposition satisfies the self-contradiction axiom. They then give the following counterexample to the converse implication.

Example 3.26 Let $\mathcal{L} = \{a, \neg a, b, \neg b\}$ and $\mathcal{R}_d = \emptyset$ and $\mathcal{R}_s = \{a \rightarrow b\} \cup \{x, \neg x \rightarrow y \mid x \in \{a, b\} \text{ and } y \in \mathcal{L}\}$. This system satisfies the self-contradiction axiom but is not closed under transposition.

It is worth noting that [Grooters, 2014] generalised all these results to the case with arbitrary logical languages with symmetric negation, c-consistent nonempty knowledge bases and reasonable argument orderings, and for both *SAF*s and for c-*SAF*s. Moreover, she did so alternatively for closure under transposition and closure under contraposition. In doing so, it was shown that the following weaker version of the self-contradiction axiom suffices:

Definition 3.27 *[Weak self-contradiction axiom] A rule-based* ASPIC$^+$ *(c-)SAF is said to satisfy the* weak self-contradiction axiom *if for each minimal inconsistent set $X \subseteq \mathcal{L}$ there is a $\sigma \in X$ such that $\neg\sigma \in Cl_{\mathcal{R}_s}(X)$.*

3.4 On the need for the various elements of *ASPIC*$^+$

ASPIC$^+$ as a general framework is quite expressive. The question therefore arises whether all these elements are really needed.

3.4.1 The need for knowledge bases

The *ASPIC* system as presented in [Caminada and Amgoud, 2007] did not have knowledge bases. Instead, certain and uncertain premises were encoded as strict rules $\rightarrow \varphi$ and defeasible rules $\Rightarrow \varphi$. Others, such as [Dung and Thang, 2014], [Li and Parsons, 2015] and [Dung, 2016] also adopt this idea. Yet there are good reasons to retain knowledge bases. To start with, the distinction between knowledge (or beliefs) and inference rules is a natural one, widely adopted in logic. Furthermore, this distinction allows a systematic study of encodings of logical consequence notions in the set of strict rules, as we will see below. We therefore conclude that although dispensing with knowledge bases might have practical advantages in specific applications, a general theory of argumentation-based inference should retain the formal distinction between knowledge and inference rules.

3.4.2 The need for strict rules and axiom premises

[Li and Parsons, 2015] show that every *ASPIC*$^+$ *SAF* with a weakest-link ordering that satisfies the rationality postulates can be translated into a *SAF* with no strict rules and no axiom premises and that (for all of [Dung, 1995]'s semantics) validates exactly the same conclusions as the original SAF. Their basic idea is that each strict rule is translated to a corresponding defeasible rule and each axiom premise to an ordinary premise, and the argument ordering is then extended so as to give the new elements resulting from the translations of strict rules or axiom premises, precedence over all conflicting elements. While this result is theoretically interesting, we still believe that the distinction between strict and defeasible inference rules is a natural one and is philosophically grounded. For example, the observation that the inclusion of strict rules allows a systematic study of encodings of logical consequence notions also applies here. We also believe that the distinction between disputable (ordinary) and undisputable (axiom) premises is a practically useful one. For these reasons we claim that a general framework for structured argumentation should leave room for these distinctions.

3.4.3 The need for preferences

In the context of ABA, [Kowalski and Toni, 1996] proposed a way to encode preferences with a specific use of assumptions in strict rules with the effect that if a preferred rule applies, the assumption in a non-preferred conflicting rule is attacked. The same can in fact be done with defeasible rules. However, [Kowalski and Toni, 1996]'s proposal does not cover any of the argument orderings discussed in this chapter. Outside of argumentation, a systematic treatment for [Brewka, 1994b; Brewka, 1994a]'s prioritised default logic was given by [Delgrande and Schaub, 2000], who showed that prioritised default theories can be translated into equivalent ordinary default theories. In Section 4.5 we will discuss the relation between prioritised default logic and $ASPIC^+$.

In general, the question as to whether $ASPIC^+$ argument orderings can be encoded in $ASPIC^+$ rule sets or knowledge bases is still an open question. We conjecture that such translations may be very hard to give for argument orderings that depend on global properties of an argument, such as weakest-link orderings.

3.4.4 The need for defeasible rules

Perhaps the most controversial issue is whether defeasible inference rules are needed. In Section 2.1 we illustrated with an informal example that there are three ways to attack an argument: on its premises, on its defeasible inferences, and on the conclusions of its defeasible inferences. In Section 2.2.2 we saw that $ASPIC^+$ explicitly allows all three forms of attack. 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 in our example of Section 2.1 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? These informal arguments have some formal backing since, as we will discuss in more detail in Section 5.2, [Dung and Thang, 2014] have shown that defeasible inference rules can in $ASPIC^+$ be reduced to strict rules.

In answer to these questions, we first 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.

We next note that some have argued that such deductive simulations are prone to yielding counterintuitive results. To illustrate, consider a instantiation of $ASPIC^+$ with no defeasible rules and in which the strict rules correspond to classical propositional logic as defined in Section 3.1.2, and assume that natural-language generalisations 'If P then normally Q' are formalised as material implications $P \supset Q$ 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: 'This alarm in this building usually does not give false alarms', so (strictly) 'false alarms in this building are usually not given by this alarm'. This strikes some as counterintuitive, since the first generalisation is consistent with the situation that this alarm is the only one in the building that gives false alarms, so the contraposition of 'If P then normally Q' cannot be deductively valid.

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 we still have the controversial deductive validity of contraposition for generalisations. In the false-alarm example the contraposition of the rule with the added normality condition would read: 'any false alarm in this building which is usual with respect to false alarms in this building cannot be this alarm', which is clearly not deductively entailed by the initial generalisation given that it is consistent with the situation that this alarm is the only one in the building that gives false alarms.

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 $\neg Q \supset \neg P$.

So far we have argued that contrapositive inferences with defeasible conditionals cannot be deductively valid (for a more detailed argument see [Modgil and Prakken, 2014, Section 4.5]). One way to respect this is to formalise defeasible natural-language conditionals as domain-specific defeasible inference rules in $ASPIC^+$ (see Section 3.1.3 above and in more detail [Modgil and Prakken, 2014]). However, this makes it hard to capture some logical properties of defeasible conditionals. For example, it might be argued that modus tollens and contraposition, although deductively invalid, are still defeasibly valid. For instance, in crime investigations the police often reason: if this person was at the crime, then we must be able to find his DNA at the crime scene; we have not been able to find his DNA at the crime scene, so presumably he was not at the crime scene. This seems a perfectly rational way of reasoning, provided that the modus-tollens inference is regarded as defeasible. Perhaps this can be captured by formalising generalisations with a defeasible object level connective $\rightsquigarrow$, as discussed above in Section 3.1.3 and by adding the appropriate strict and defeasible inference rules for $\rightsquigarrow$ to $\mathcal{R}_s$ and $\mathcal{R}_d$. For example, defeasible modus tollens could be added as follows:

$$\neg\psi, \varphi \rightsquigarrow \psi \Rightarrow \neg\varphi$$

However, doing so is not straightforward, since the above encoding of the defeasible modus pollens principle is in the form of an inference rule used in construction of $ASPIC^+$ arguments, while in contrast, the current nonmonotonic logics for defeasible conditionals model such principles at the level of the consequence relation (which in $ASPIC^+$ is defined in terms of the outcome of argument evaluation; cf. Definition 2.18 above). This suggests the following topic for future research: how to instantiate the sets of strict and defeasible rules in $ASPIC^+$ in such a way that the semantic insights on defeasible conditionals obtained in other areas of nonmonotonic logic are respected?

So far our discussion has focused on argumentation based reasoning as it applies to beliefs (i.e., reasoning about what is the case, often called *epistemic reasoning* by philosophers). However 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, since statistics only applies to epistemic reasoning. Space limitations prevent us from giving more details about these philosophical arguments.

Finally, as further discussed in Section 4.1, [Dung and Thang, 2014] show for the case without preferences and knowledge bases that $ASPIC^+$ defeasible rules can be equivalently translated into theories of assumption-based argumentation (ABA). Since, as also discussed further in Section 4.1, ABA can be reconstructed as a special case of $ASPIC^+$ with no knowledge bases, defeasible rules or preferences, [Dung and Thang, 2014]'s result implies that the defeasible rules of $ASPIC^+$ SAFs with no knowledge bases or preferences can be translated into strict $ASPIC^+$ rules.

3.4.5 The value of translation results

Translation results like the ones of [Dung and Thang, 2014] and [Li and Parsons, 2015] on translating one type of rule into the other, and possible future results on encoding preferences in rules, are theoretically interesting and may have practical benefits. For example, [Dung and Thang, 2014]'s result makes it possible to use ABA tools for implementing fragments of $ASPIC^+$ without preferences. However, such translation results should be interpreted with care. Logic is full of such results and they do not necessarily mean that the translated system is less useful or less interesting. For example, nobody would say that the fact that all connectives of propositional logic can be translated into a single one means that presentations of propositional logic with the usual five or six connectives are unnecessarily complicated; on the contrary, versions with just one connective would lead to unnecessarily complex knowledge representations. Likewise, versions of $ASPIC^+$ with both strict and defeasible rules and with preferences may lead to more compact and more natural representations. Moreover, nobody would say that translations of modal logic into first-order predicate logic show that modal logic is superfluous. On the contrary, modal logics often provide systematic treatments of modalities in ways that their first-order translations do not. Likewise, $ASPIC^+$ provides a theory of reasoning with a combination of strict and defeasible rules and allows a general study of argumentation with preferences, something which formalisms with only strict or only defeasible rules or formalisms without preferences do not provide.

3.5 Argument schemes and critical questions

We concluded Section 3.1.3 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 $ASPIC^+$ and how strict inference rules can also be accommodated when doing so.

John Pollock formalised defeasible rules for reasoning patterns involving perception, memory, induction, temporal persistence and the statistical syllogism, as well as undercutters for these reasons. In $ASPIC^+$ his principles of perception and memory can be written as follows:

$$\begin{array}{ll} d_p(x,\varphi): & \texttt{Sees}(x,\varphi) \Rightarrow \varphi \\ d_m(x,\varphi): & \texttt{Recalls}(x,\varphi) \Rightarrow \varphi \end{array}$$

In fact, these defeasible inference rules are schemes for all their ground instances (that is, for any instance where x and φ 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 φ 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

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

combined with information such as

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

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 (trustworty, 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 question (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:

$d_w(x,\varphi)$: $\texttt{PositionToKnow}(x,\varphi), \texttt{Says}(x,\varphi) \Rightarrow \varphi$
$u_w(x,\varphi)$: $\neg\texttt{Credible}(x) \Rightarrow \neg d_w(x,\varphi)$

We will now illustrate the modelling of both Pollock's defeasible reasons and Walton's argument schemes with our example from Section 2.1, 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, that Jan testified that he saw John in Amsterdam on the same morning, and that 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: $\texttt{PositionToKnow}(Bob, \texttt{Recalls}(Bob, \texttt{Sees}(Bob, \texttt{InHollandPark}(John))))$
f_2: $\texttt{Says}(Bob, \texttt{Recalls}(Bob, \texttt{Sees}(Bob, \texttt{InHollandPark}(John))))$
f_3: $\texttt{SeesPicturesAbout}(Bob, \texttt{Sees}(Bob, \texttt{InHollandPark}(John)))$
f_4: $\forall x,\varphi.(\texttt{SeesPicturesAbout}(x,\varphi) \supset \texttt{DistortedMemory}(x,\varphi))$
f_5: $\forall x.\texttt{InHollandPark}(x) \supset \texttt{InLondon}(x)$
f_6: $\texttt{PositionToKnow}(Jan, \texttt{Recalls}(Jan, \texttt{Sees}(Jan, \texttt{InAmsterdam}(John))))$
f_7: $\texttt{Says}(Jan, \texttt{Recalls}(Jan, \texttt{Sees}(Jan, \texttt{InAmsterdam}(John))))$
f_8: $\texttt{Friends}(Jan, John)$
f_9: $\texttt{SuspectedRobber}(John)$
f_{10}: $\forall x,y,\varphi.\texttt{Friends}(x,y) \wedge \texttt{SuspectedRobber}(y) \wedge \texttt{InvolvedIn}(y,\varphi) \supset \neg\texttt{Credible}(x)$
f_{11}: $\texttt{InvolvedIn}(John, \texttt{Recalls}(Jan, \texttt{Sees}(Jan, \texttt{InAmsterdam}(John))))$
f_{12}: $\forall x\neg(\texttt{InAmsterdam}(x) \wedge \texttt{InLondon}(x))$

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

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

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

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

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

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

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

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

A_8: $A_7, f_5, f_{12} \Rightarrow \neg\texttt{InAmsterdam}(John)$

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 us 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 Relationship with other Argumentation Formalisms

As shown in various publications on *ASPIC*$^+$, its generality allows the reconstruction of various other systems and frameworks as special cases of *ASPIC*$^+$. In this section we review this work in some detail. We also discuss the relationship of *ASPIC*$^+$ with various developments of abstract argumentation frameworks.

4.1 Assumption-based argumentation

Assumption-based argumentation (ABA) emerged from attempts to give an argumentation-theoretic semantics to logic-programming's negation as failure, and has developed into a general framework for nonmonotonic logics [Bondarenko *et al.*, 1993; Bondarenko *et al.*, 1997; Toni, 2014]. ABA assumes a 'deductive system', consisting of a set of inference rules defined over some logical language. Given a set of so-called 'assumptions' formulated in the logical language, arguments are then deductions of claims using rules and supported by sets of assumptions. In general, ABA leaves both the logical language and set of inference rules unspecified, so that like $ASPIC^+$, it is an abstract framework for structured argumentation. However, unlike $ASPIC^+$, ABA only allows attacks on an argument's assumptions, so that ABA's rules are effectively strict inference rules. In order to express conflicts between arguments, ABA makes a minimum assumption on the logical language, namely, that each assumption has a contrary. That b is a contrary of a, written as $b = \overline{a}$, informally means that b contradicts a. An argument using an assumption a is then attacked by any argument for conclusion $\overline{a}$. In [Bondarenko *et al.*, 1997] an argumentation-theoretic semantics is then given which is very much like [Dung, 1995]'s abstract approach, except that [Bondarenko *et al.*, 1997] considers sets of assumptions rather than sets of arguments. However, [Dung *et al.*, 2007] showed that an equivalent fully argument-based formulation can be given.

In this section we first discuss how ABA can be reconstructed in $ASPIC^+$ and then how some instantiations of $ASPIC^+$ can be reconstructed in ABA.

4.1.1 Reconstructing ABA in $ASPIC^+$

Above we remarked that [Bondarenko *et al.*, 1997]'s version of ABA is strictly speaking not an instantiation of [Dung, 1995]'s abstract argumentation frameworks but that [Dung *et al.*, 2007] gave an equivalent formulation of ABA in such frameworks. [Prakken, 2010] showed that this reconstructed version of ABA can in turn be reconstructed as a special case of $ASPIC^+$ extended with possibly non-symmetric negation (see Section 2.3 above). In $ASPIC^+$ as defined by [Prakken, 2010], the ordinary premises were further divided into 'really' ordinary premises and assumptions and the assumption premises were used to model ABA assumptions. However, as observed by [Modgil and Prakken, 2013, Section 3.1], one can do without such specialised premises and model assumptions as ordinary premises. ABA can then be reconstructed as the special case of $ASPIC^+$ with empty sets of defeasible rules and axiom premises and no preferences.

First the main definitions of ABA are recalled.

Definition 4.1 *(Def. 2.3 of [Dung* et al.*, 2007].) A* deductive system *is a pair* $(\mathcal{L}, \mathcal{R})$ *where*

- $\mathcal{L}$ *is a formal language consisting of countably many sentences, and*

- $\mathcal{R}$ *is a countable set of inference rules of the form* $\alpha_1, \ldots, \alpha_n \to \alpha$.[13] $\alpha \in \mathcal{L}$ *is called the* conclusion *of the inference rule,* $\alpha_1, \ldots, \alpha_n \in \mathcal{L}$ *are called the* premises *of the inference rule and* $n \geq 0$.

Definition 4.2 *(Def. 2.5 of [Dung* et al.*, 2007].) An* assumption-based argumentation framework *(ABF) is a tuple* $(\mathcal{L}, \mathcal{R}, \mathcal{A}, ^-)$ *where*

- $(\mathcal{L}, \mathcal{R})$ *is a deductive system.*
- $\mathcal{A} \subseteq \mathcal{L}$, $\mathcal{A} \neq \emptyset$. $\mathcal{A}$ *is the set of* assumptions.
- *If* $\alpha \in \mathcal{A}$, *then there is no inference rule of the form* $\alpha_1, \ldots, \alpha_n \to \alpha \in \mathcal{R}$.
- $^-$ *is a total mapping from* $\mathcal{A}$ *into* $\mathcal{L}$. $\overline{\alpha}$ *is the* contrary *of* α.

ABA arguments are then defined in terms of deductions. To remain as close as possible to *ASPIC*$^+$, we here give the tree-based definition of [Toni, 2014] (with some minor stylistic rephrasings). The proofs of [Prakken, 2010] instead use [Dung *et al.*, 2007]'s sequence-based definition, which essentially presents one particular order in which a tree-style argument can be constructed.

Definition 4.3 *([Toni, 2014].) A* deduction *for a conclusion* α *supported by premises* $S \subseteq \mathcal{L}$ *is a finite tree with nodes labelled by sentences in* $\mathcal{L}$ *or by* τ[14]. *Each leaf is either* τ *or a sentence in* S. *each non-leave* α' *has, as children, the elements of the body of some rule in* $\mathcal{R}$ *with head* α'.

Then an assumption-based argument is defined as follows.

Definition 4.4 *(Def. 2.6 of [Dung* et al.*, 2007].) An* argument *for a conclusion on the basis of an ABF is a deduction of that conclusion whose premises are all assumptions (in* $\mathcal{A}$*).*

As for notation, the existence of an argument for a conclusion α supported by a set of assumptions A is denoted by $A \vdash \alpha$, or by $A \vdash_{ABF} \alpha$ if it has to be distinguished from the existence of a strict argument according to Definition 2.4 with the same premises and conclusion; the latter will below be denoted by $A \vdash_{AT} \alpha$.

Finally, Dung et al.'s notion of argument attack is defined as follows.

Definition 4.5 *(Def. 2.7 of [Dung* et al.*, 2007].)*

- *An argument* $A \vdash \alpha$ attacks an argument $B \vdash \beta$ *if and only if* $A \vdash \alpha$ *attacks an assumption in* B;
- *an argument* $A \vdash \alpha$ attacks an assumption β *if and only if* α *is the contrary* $\overline{\beta}$ *of* β.

[13] In [Dung *et al.*, 2007] the arrows are from right to left.
[14] τ represents 'true' and stands for the empty body of rules.

The $ASPIC^+$ argumentation theory corresponding to an assumption-based argumentation framework is then in [Prakken, 2010] defined as follows.[15]

Definition 4.6 *[Mapping ABFs to ATs] Given an assumption-based argumentation framework* $ABF = (\mathcal{L}_{ABF}, \mathcal{R}_{ABF}, \mathcal{A}, {}^{-}{}_{ABF})$*, the corresponding argumentation theory* $AT_{ABF} = (AS, \mathcal{K})$*, where* $AS = (\mathcal{L}_{AT}, {}^{-}{}_{AT}, \mathcal{R}_{AT}, n)$ *and* $\mathcal{K} = K_n \cup \mathcal{K}_p$*, is defined as follows:*

- $\mathcal{L}_{AT} = \mathcal{L}_{ABF}$
- $\varphi \in \overline{\psi}_{AT}$ *iff* $\varphi = \overline{\psi}_{ABF}$
- $\mathcal{R}_{AT} = \mathcal{R}_s = \mathcal{R}_{ABF}$
- $\mathcal{K}_n = \emptyset$
- $\mathcal{K}_p = \mathcal{A}$
- n *is undefined.*

Then it can be shown that for all ABFs: there exists an argument $A \vdash_{ABF} \alpha$ if and only if there exists an argument $A \vdash_{AT} \alpha$. From this it follows for all ABFs and for every argument $A \vdash_{ABF} \alpha$ and every argument $A \vdash_{AT} \alpha$: $A \vdash_{ABF} \alpha$ is attacked by an argument $B \vdash_{ABF} \beta$ if and only if $A \vdash_{AT} \alpha$ is defeated by an argument $B \vdash_{AT} \beta$. Then the main correspondence result can be proven:

Theorem 4.7 (Thm. 8.8 of [Prakken, 2010]) *For all ABFs, and for any semantics S subsumed by complete semantics and any set E:*

1. *if E is an S-extension of ABF then E_{AT} is an S-extension of AT, where* $E_{AT} = \{A \vdash_{AT} \alpha \mid A \vdash_{ABF} \alpha \in E\}$*;*
2. *if E is an S-extension of AT then E_{ABF} is an S-extension of ABF, where* $E_{ABF} = \{A \vdash_{ABF} \alpha \mid A \vdash_{AT} \alpha \in E\}$*.*

Theorem 4.7 says that there is a one-to-one correspondence between the extensions of an ABF and those of its corresponding AT. Note also that the above results carry over to [Verheij, 2003]'s DefLog argumentation system since, as observed by Verheij, DefLog can be translated into ABA.

One virtue of this reconstruction of ABA in $ASPIC^+$ is that one can then identify conditions under which ABA satisfies rationality postulates (by requiring, for instance, that the strict rules are closed under transposition).

[15] In fact, in [Prakken, 2010] the ABA assumptions were translated into $ASPIC^+$ assumption-type premises, which in [Prakken, 2010] was an additional category of premises. However, as remarked by [Modgil and Prakken, 2013], the translation also succeeds when defined as below.

4.1.2 Reconstructing instantiations of *ASPIC*$^+$ in ABA

[Dung and Thang, 2014] have shown that their rule-based systems, which are a special case of $ASPIC^+$ with no knowledge base and no preferences, can be translated into ABA instantiations. They do this by translating every defeasible rule $p_1, \ldots, p_n \Rightarrow q$ as a strict rule $d_i, p_1, \ldots, p_n, not\neg q \to q$, where

- $d_i = n(p_1, \ldots, p_n \Rightarrow q)$ in $ASPIC^+$;
- $d_i, not\neg q \in \mathcal{A}$ (i.e., they are ABA assumptions);
- $q = \overline{not\neg q}$ and for all φ: $\varphi = \overline{\neg\varphi}$ and $\neg\varphi = \overline{\varphi}$

[Dung and Thang, 2014] then show (on the assumption that $ASPIC^+$ rule names do not occur as antecedents or consequents in $ASPIC^+$ rules), that for each semantics subsumed by complete semantics the resulting ABA framework validates the same conclusions as the original $ASPIC^+$ SAF. Generalising this result to cases with preferences is still an open question.

4.2 Tarskian abstract logics and classical-logic argumentation

[Amgoud and Besnard, 2013] present an abstract approach to defining the structure of arguments and attacks, based on Tarski's notion of an abstract logic that only assumes 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 (see [Amgoud and Besnard, 2013] for a more detailed account of these constraints). Finally, Tarski defined a set $S \subseteq \mathcal{L}$ as *consistent* iff $Cn(S) \neq \mathcal{L}$. In [Amgoud and Besnard, 2013], an argument is a pair (S, p) where $S \subseteq \mathcal{L}$ is consistent, $p \in Cn(S)$ and S is a minimal (under set inclusion) set satisfying these conditions. Then (S, p) attacks (T, q) iff $\{p, q'\}$ is inconsistent for some $q' \in T$.

[Modgil and Prakken, 2013, Section 5.2] show that $ASPIC^+$ can be used to reconstruct, and extend with preferences, the Tarskian logic approach. For the strict rules, they choose (for any finite $S \subseteq \mathcal{L}$):

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

Then given any $\Sigma \subseteq \mathcal{L}$, they let $\mathcal{K}_p = \Sigma$, $\mathcal{R}_d = \emptyset$. Also, $\forall\phi \in \mathcal{L}$, ϕ has a contradictory ψ, and if $\phi = -\psi$ then $Cn(\{\phi, \psi\}) = \mathcal{L}$ and if $Cn(\{\phi, \psi\}) = \mathcal{L}$ then $\exists\phi' \in Cn(\{\phi\})$ s.t. $\phi' = -\psi$. They then show that given a reasonable argument preference ordering $\preceq$ (possibly defined on the basis of an ordering $\leq'$ over Σ), the c-SAF is well defined. Hence one obtains an account of [Amgoud and Besnard, 2013]'s Tarskian logic abstract argumentation approach that is extended with preferences and is well behaved with respect to rationality postulates. Two issues to note are that the reconstruction employs $ASPIC^+$ undermining attacks, which differ from the abstract logic attacks defined above which rely on showing that the claim and attacked premises are inconsistent. However, [Modgil and Prakken, 2013] show that the use of $ASPIC^+$ attacks does not change the outcome in the sense that the complete (and hence grounded, preferred and stable) extensions remain the same irrespective of whether we use the abstract logic notion of an attack instead. Moreover, *ASPIC*$^+$ imposes no subset minimality conditions on the premises of arguments. However, [Modgil and Prakken, 2013] show that if subset minimal arguments are not strengthened by adding 'irrelevant' premises – i.e., if A is subset minimal and $A \not\prec B$ then $A' \not\prec B$ where $\texttt{Prem}(A') \supset \texttt{Prem}(A)$ – then the conclusions of arguments in complete extensions remains the same whether or not we exclude arguments that are not subset minimal.

[Modgil and Prakken, 2013] then applied this to a reconstruction of so-called classical argumentation [Cayrol, 1995; Besnard and Hunter, 2001; Besnard and Hunter, 2008; Gorogiannis and Hunter, 2011], which formalises arguments as minimal classical consequences from consistent and finite premise sets in standard propositional or first-order logic. In particular, [Gorogiannis and Hunter, 2011] study classical logic instantiations of abstract argumentation frameworks. [Modgil and Prakken, 2013] reconstruct this as a specific instance of the above formulation of the Tarskian abstract logic approach, with Cn the classical consequence operator (below denoted as $\models$). This yields the following instantiation of *ASPIC*$^+$:

Definition 4.8 *[Classical argumentation with preferences reconstructed in* ASPIC$^+$*]* *Let $\mathcal{L}'$ be a classical-logic language, $\Sigma \subseteq \mathcal{L}'$ and $\leq'$ a partial preorder on Σ. A* classical-logic argumentation theory *based on $(\mathcal{L}', \Sigma, \leq')$ is a pair $(AS, \mathcal{K})$ such that AS is an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, n)$ where:*

1. *$\mathcal{L} = \mathcal{L}'$;*
2. *$\varphi \in \overline{\psi}$ iff $\varphi = \neg\psi$ or $\psi = \neg\varphi$;*
3. *$\mathcal{R}_d = \emptyset$, and for all finite $S \subseteq \mathcal{L}$ and $p \in \mathcal{L}$, $S \to p \in \mathcal{R}_s$ iff $S \models p$.*

$\mathcal{K}$ is a knowledge base such that $\mathcal{K}_n = \emptyset$ and $\mathcal{K}_p = \Sigma$.
$(\mathcal{A}, \mathcal{C}, \preceq)$ is the c-SAF *based on $(AS, \mathcal{K})$ as defined in Definition 2.14 and where $\preceq$ is defined in terms of $\leq'$ as in Section 3.2.*

[Gorogiannis and Hunter, 2011] define seven attack relations and prove that only the following two ensure satisfaction of the rationality postulate of indirect consistency:

- *Y directly undercuts X* if $\texttt{Conc}(Y) \equiv \neg p$ for some $p \in \texttt{Prem}(X)$
- *Y directly defeats X* if $\texttt{Conc}(Y) \vdash_c \neg p$ for some $p \in \texttt{Prem}(X)$

Since classical logic can be specified as a Tarskian abstract logic, [Modgil and Prakken, 2013] can prove via their reconstruction of abstract-logic argumentation, that the *ASPIC*$^+$ notion of undermining attacks is equivalent to direct undercuts and defeats in that the complete extensions generated are the same. Moreover, from the results described above in Section 3.2 it follows that their extension of classical-logic argumentation with preferences satisfies the rationality postulates. Indeed, [Modgil and Prakken, 2013] argue that the extension to include preferences is needed if classical-logic argumentation is to be effectively used in arbitrating amongst conflicts, since as shown in ([Cayrol, 1995; Gorogiannis and Hunter, 2011; Amgoud and Besnard, 2013]), there is a one-to-one correspondence between the (premises of arguments in in) preferred/stable extensions of abstract argumentation frameworks instantiated by a classical-logic knowledge base and the maximal consistent subsets of the knowledge base. This is to be expected, given the monotonicity of classical logic (and thus the absence of logical mechanisms to withdraw previously derivable contradictory inferences).

4.3 Carneades

As shown by [Van Gijzel and Prakken, 2011; Van Gijzel and Prakken, 2012], the Carneades system of [Gordon *et al.*, 2007; Gordon and Walton, 2009b] can be reconstructed as a special case of basic $ASPIC^+$ with a generalised contrariness relation. A Carneades argument is a triple $\langle P, E, c\rangle$ where P is a set of *premises*, E a set of *exceptions* and c the *conclusion*, which is either pro or con a *statement* s. Carneades does not assume that premises and conclusions are connected by inference rules. Also, all arguments are elementary, that is, they contain a single inference step; they are combined in recursive definitions of *applicability* of an argument and *acceptability* of its conclusion. In essence, an *argument* is *applicable* if (1) all its premises are given as facts or else are acceptable conclusions of other arguments, and (2) none of its exceptions are given as facts or as acceptable conclusions of other arguments. A *statement* is *acceptable* if it satisfies its *proof standard*. Facts are stated by an *audience*, which also provides numerical *weights* for each argument plus *thresholds* for argument weights and differences in argument weights. In the publications on Carneades five proof standards are defined. One is *preponderance of the evidence*:

> Statement p satisfies *preponderance of the evidence* iff there exists at least one applicable argument pro p for which the weight is greater than the weight of any applicable argument con p.

In the $ASPIC^+$ reconstruction of Carneades the facts are reconstructed as elements of $\mathcal{K}_n$, while the Carneades notions of applicability and acceptability are encoded in the $ASPIC^+$ defeasible inference rules. For every Carneades argument $a = \langle P, E, c\rangle$, a defeasible rule $P \Rightarrow_{app_a} arg_a$ is added, saying that if P then a is applicable[16]. Moreover, a defeasible rule $arg_a \Rightarrow_{acc_a} c$ is added, saying that if a is applicable, its conclusion is acceptable. Here, app_a and acc_a are the respective names of these rules in $\mathcal{L}$ according to the naming convention n. Thus a Carneades argument $\langle P, E, c\rangle$ pro statement s induces an $ASPIC^+$ argument:

[16] The idea to make the applicability step explicit by means of an argument node was adapted from [Brewka and Gordon, 2010].

A_1: P
A_2: $A_1 \Rightarrow_{app_a} arg_a$
A_3: $A_2 \Rightarrow_{acc_a} s$

It should be noted that effectively, a Carneades argument is analogous to a defeasible inference rule, since the representation (P, E, c) does not assume that the facts P are given as part of the argument; rather it is the *applicability* of the argument that depends on facts or arguments for P. This justifies the translation of Carneades arguments into $ASPIC^+$ defeasible rules.

Next, for each exception $e \in E$, a rule $e \Rightarrow \neg app_a$ is added to $\mathcal{R}_d$ and $\neg app_a = \overline{app_a}$ is added to the contrariness relation. So such rules can be used to undercut the $ASPIC^+$ version of an argument on its first step. Moreover, for each argument b with a conclusion c' that conflicts with s, we have that $arg_b = \overline{acc_a}$ if this is dictated by the proof standard for s. For example, if the standard for s is preponderance of the evidence, then $arg_b = \overline{acc_a}$ just in case $weight(a) \leq weight(b)$. Thus the Carneades proof standards and argument weights are not incorporated in the $ASPIC^+$ argument ordering but in the $ASPIC^+$ contrariness relation.

For example, a Carneades argument $b = \langle P', E', c' \rangle$ where c' is con s, induces an $ASPIC^+$ argument:

B_1: P'
B_2: $B_1 \Rightarrow_{app_b} arg_b$
B_3: $B_2 \Rightarrow_{acc_b} \neg s$

Then A_3 rebuts B_3 if $weight(b) < weight(a)$, B_3 rebuts A_3 if $weight(a) < weight(b)$ and both rebut each other if $weight(a) = weight(b)$. Since in the $ASPIC^+$ reconstruction all defeasible arguments are equally strong, all these rebutting attacks succeed as defeat.

[Van Gijzel and Prakken, 2011; Van Gijzel and Prakken, 2012] then prove that under this reconstruction, $ASPIC^+$ SAFs corresponding to a Carneades theory always have a unique extension, which is the same in all of [Dung, 1995]'s semantics. This perhaps surprising result is partly due to strong non-circularity assumptions made in Carneades on its 'inference graph', which contains all constructible arguments. [Van Gijzel and Prakken, 2011; Van Gijzel and Prakken, 2012] also prove that the conclusions of the justified arguments in $ASPIC^+$ correspond to the conclusions of the acceptable arguments in Carneades.

4.4 Defeasible Logic Programming

Defeasible logic programming (DeLP) is a logic-programming-based argumentation system originating from (but not equivalent to) [Simari and Loui, 1992]. The main publication on DeLP is [Garcia and Simari, 2004], which we will take as the basis for our discussion. Although DeLP is similar to $ASPIC^+$, it cannot be fully reconstructed as an instance. Elements of DeLP that instantiate $ASPIC^+$ are a predicate-logic literal language with ordinary negation, a set of indisputable facts, two sets of strict and defeasible rules, and a binary argument ordering. DeLP arguments can be reconstructed as $ASPIC^+$ arguments with the additional constraint that their sets of conclusions are consistent under application of strict rules in that for no φ it holds that $\texttt{Conc}(A) \vdash \varphi, \neg\varphi$.

DeLP's definition of attack is similar but not equivalent to $ASPIC^+$'s notion of rebutting attack. Instead (and translated to $ASPIC^+$ vocabulary), A attacks B at B's subargument B' if $\texttt{Conc}(A) \cup \texttt{Conc}(B') \vdash \varphi, \neg\varphi$ for some wff φ. Note that this allows an attack on a conclusion of a strict rule, but such an attack will never exist without an attack on a previous defeasible step in the argument as well. Apart from this difference, DeLP's notion of defeat is defined as in $ASPIC^+$: A defeats B if A attacks B on B' and $A \nprec B'$. It remains to be investigated whether adopting DeLP's notion of rebutting attack in $ASPIC^+$ would lead to different outcomes.

A main difference with $ASPIC^+$ is that DeLP as defined in [Garcia and Simari, 2004] does not evaluate arguments by generating abstract argumentation frameworks. Instead, DeLP's notion of *warrant* is defined in a way that is similar to the argument game of grounded semantics [Prakken, 1999; Modgil and Caminada, 2009] but with some significant differences. Briefly, the argument game for grounded semantics is between a proponent and an opponent of an argument A, where the proponent begins with A and then the players take turns such that the opponent defeats or strictly defeats the proponent's previous argument while the proponent strictly defeats the opponent's previous argument; in addition, the proponent is not allowed to repeat his own arguments. An argument A is justified if the proponent has a winning strategy in a game starting with A. DeLP's notion of warrant is equivalent to this notion of justification but its game rules are different. First, if one player weakly defeats the previous argument then the next player must strictly defeat that argument, while if one player strictly defeats the previous argument then the next player may either weakly or strictly defeat it. Second, no player may reuse a subargument from one of its earlier moves.

It would be interesting to adopt the game rules of grounded semantics in DeLP's notion of warrant, which would then establish a clear link between DeLP and the theory of abstract argumentation. Among other things, this would facilitate the study of the satisfaction of rationality postulates in DeLP.

4.5 $ASPIC^+$ characterisations of non-monotonic reasoning formalisms

A key reason for the prominence of argumentation (in particular Dung's theory of abstract argumentation frameworks) in knowledge representation and reasoning, is its characterisation of non-monotonic reasoning in terms of the dialectical exchange of argument and counter-argument. Indeed, in [Dung, 1995], argumentation-based characterisations of logic programming, [Reiter, 1980]'s Default Logic and [Pollock, 1987]'s argumentation system are formalised. The theory thus provides foundations for reasoning by individual computational and human agents, and distributed non-monotonic reasoning ('dialogue') amongst agents.

ASPIC$^+$ continues in this tradition, formalising logic programming instantiations of abstract argumentation frameworks, whereby the defeasible rules are rules in a logic program, the strict rules and axiom premises are empty, the preference relation is empty, and (as described in Section 2.3) the ordinary premises are the negation as failure ($\sim$) assumptions in the antecedents of defeasible rules, and we define the contrary function $\forall \alpha \in \mathcal{L}$: $\alpha \in \overline{\sim \alpha}$.

Brewka's Preferred Subtheories [Brewka, 1989] can also be formalised as an instance of *ASPIC*$^+$'s formalisation of classical-logic argumentation (as outlined in Section 4.2). The arguments and attacks are defined by a base Σ of propositional classical wff equipped with a total ordering $\leq'$ which is used by the set comparison $\triangleleft_{\texttt{Eli}}$ to define weakest link preferences over arguments. One then obtains an argumentation-based characterisation of non-monotonic inference defined by Preferred Subtheories. The latter starts with a stratification $(\Sigma_1, \ldots, \Sigma_n)$ of the totally ordered Σ ($\alpha, \beta \in \Sigma_i$ iff $\alpha \equiv' \beta$ and $\alpha \in \Sigma_i, \beta \in \Sigma_j, i < j$ iff $\beta \in \Sigma <' \alpha \in \Sigma$). A 'preferred subtheory' (*ps*) is obtained by taking a maximal under set inclusion consistent subset of Σ_1, maximally extending this with a subset of Σ_2, and so on. Multiple individually consistent preferred subtheories may be constructed, and [Modgil and Prakken, 2013] show that each *ps* corresponds to the premises of arguments in a stable extension. Hence, α is classically entailed from a *ps* iff α is the conclusion of an argument in a stable extension. Then α is a sceptical (credulous) Preferred Subtheories inference iff α is entailed by all (respectively at least one) *ps*, iff α is sceptically (credulously) justified under the stable semantics (as defined in Definition 2.18).

More recently, *ASPIC*$^+$ has been used to provide an argumentative characterisation of Brewka's *Prioritised Default Logic* (*PDL*) [Brewka, 1994a]. *PDL* upgrades [Reiter, 1980]'s Default Logic to include a strict partial ordering $<_D$ on a finite set D of first order normal defaults of the form $\frac{\theta:\phi}{\phi}$. Then given a set W of first order formulae, and a linearisation $<^+$ of $<_D$, one iteratively applies the highest ordered default whose antecedent is in the first order closure of the result obtained in the previous iteration. Intuitively, one starts with the classical consequences E_0 of W, and then adds the consequent of the highest ordered default whose antecedent is contained in E_0. Then closure under classical consequence obtains E_1, to which one adds the consequent of the highest ordered default whose antecedent is contained in E_1, and so on, until $E_{n+1} = E_n$ is the unique extension of $(D, W, <)$. In [Young *et al.*, 2016], an *ASPIC*$^+$ *SAF* is defined in which the contrary function is defined so as to formalise classical negation, $\mathcal{R}_s$ characterises inference in first order classical logic, the axiom premises $\mathcal{K}_n$ is defined as W ($\mathcal{K}_p = \emptyset$), $\mathcal{R}_d = \{\theta \Rightarrow \phi | \frac{\theta:\phi}{\phi} \in D\}$ (with the naming function n undefined), and $<_D$ the ordering on $\mathcal{R}_d$. A linear 'structure preference' ordering $<_{SP}$ is defined, which modifies $<_D$ so as to account for the dependency amongst rules in $\mathcal{R}_d$ (i.e., for any set of rules applicable given all rules thus far applied, $<_{SP}$ picks out the $<_D$ maximal rule, and the process is repeated for the set of rules that are subsequently applicable). Then the *disjoint* elitist ordering – $\Gamma \lhd_{\texttt{DEli}} \Gamma'$ iff $\exists r \in \Gamma \setminus \Gamma', \forall r' \in \Gamma' \setminus \Gamma : r <_{SP} r'$ – is used to define an ordering over arguments according to the weakest link principle. [Young *et al.*, 2016] then show that the single extension E of $(D, W, <)$ corresponds to the conclusions of arguments in the (provably) unique stable extension of the corresponding $ASPIC^+$ SAF.

4.6 The relationship of $ASPIC^+$ with developments of the theory of abstract argumentation frameworks

$ASPIC^+$ is designed to generate abstract argumentation frameworks in the sense of [Dung, 1995]. Over the years, various extensions of abstract argumentation frameworks with further elements have been proposed, such as with preferences ([Amgoud and Cayrol, 1998]'s preference-based argumentation frameworks or $PAFs$), values ([Bench-Capon, 2003]'s value-based argumentation frameworks or $VAFs$), attacks on attacks ([Modgil, 2009]'s extended argumentation frameworks or $EAFs$) and abstract support relations between arguments (e.g. [Cayrol and Lagasquie-Schiex, 2009]'s bipolar argumentation frameworks or $BAFs$). The question arises as to what extent $ASPIC^+$ can be seen as instantiations of these frameworks. Moreover, work has recently been done on the dynamics of abstract argumentation frameworks, such as deleting or adding arguments or attacks; e.g. [Baroni and Giacomin, 2008; Baroni *et al.*, 2011b; Baumann and Brewka, 2010]. The question also arises as to what extent can the dynamics of argumentation, as studied in these works, be applied to $ASPIC^+$. These questions are answered in this section.

4.6.1 E-$ASPIC^+$: Structuring Extended Argumentation Frameworks

[Modgil, 2009] extended abstract argumentation frameworks to accommodate arguments that attack attacks, and in so doing enabled integration of arguments that express preferences over other arguments. The essential idea is that given an attack from A to B, then if the argument C expresses a strict preference for B over A, C attacks (and so invalidates the success of) the attack from A to B. A modified definition of the acceptability of arguments was defined for these *Extended Argumentation Frameworks* (EAFs), and [Modgil, 2009] showed that one can reconstruct [Prakken and Sartor, 1997]'s logic-programming-based argumentation system with defeasible preferences as an instance of EAFs. In this reconstruction, arguments built from rules expressing preferences over other 'object level' rules, constitute arguments expressing preferences over the arguments built from the object level rules.

However, as with Dung's original abstract argumentation frameworks, the abstract EAFs can in principle yield extensions that violate the rationality postulates. Hence [Modgil and Prakken, 2010] define a version of $ASPIC^+$ – E-$ASPIC^+$ – that generate a special class of *bounded hierarchical* EAFs in which the finite arguments $\mathcal{A}$ can be stratified into $\mathcal{A}_1, \ldots, \mathcal{A}_n$, such that if $C \in \mathcal{A}_i$ ($i \neq 1$) expresses a preference for B over A, then $A, B \in \mathcal{A}_{i-1}$. As in $ASPIC^+$ arguments are constructed from strict and defeasible rules, and axiom and ordinary premises, and in addition to the usual notions of attack, E-$ASPIC^+$ defines a function over sets of arguments $\mathcal{A}' \subseteq \mathcal{A}$, that maps $\mathcal{A}'$ to a strict preference over some $B, A \in \mathcal{A}$. In this way, EAFs are conservatively modified to allow for attacks on attacks to originate from *sets of*, rather than *single*, arguments. As well as the notion of a well-defined SAF[17] [Modgil and Prakken, 2010] additionally identify a condition that if $\mathcal{A}' \subseteq \mathcal{A}$ expresses that $A \prec B$ and $\mathcal{A}'' \subseteq \mathcal{A}$ expresses that $B \prec A$, then $\mathcal{A}'$ and $\mathcal{A}''$ respectively contain arguments X and Y that have contradictory conclusions, or some X and Y such that X can be extended by strict rules to an argument X^+ such that X^+ and Y have contradictory conclusions. [Modgil and Prakken, 2010] then show that the generated *bounded hierarchical* EAFs satisfy [Caminada and Amgoud, 2007]'s rationality postulates.

4.6.2 Abstract support relations

There have been several recent proposals to extend abstract argumentation frameworks with abstract support relations, such as [Cayrol and Lagasquie-Schiex, 2005; Cayrol and Lagasquie-Schiex, 2009; Cayrol and Lagasquie-Schiex, 2013]'s Bipolar Argumentation Frameworks ($BAFs$), the work of [Martinez *et al.*, 2006] and [Oren and Norman, 2008]'s Evidential Argumentation Systems ($EASs$). Various semantics for such frameworks have been defined, claiming to capture different notions of support. For example, [Boella *et al.*, 2010a] study semantics of what they call "deductive" support, which satisfies the constraint that if A is acceptable and A is a deductive support of B, then B is acceptable. [Nouioua and Risch, 2011] consider "necessary support", which satisfies the constraint that if B is acceptable and A is a necessary support of B, then A is acceptable.

One question is whether the $ASPIC^+$ notion of a subargument instantiates any of these notions. Here we first discuss [Dung and Thang, 2014]'s simple way of formalising [Nouioua and Risch, 2011] intuitions concerning necessary support, namely, by adding a binary support relation $\mathcal{S}$ on $\mathcal{A}$ to AFs with the sole additional constraint that if B supports C and A defeats B then A also defeats C. The semantics of the resulting abstract argumentation frameworks is simply defined by choosing one of the semantics for the corresponding pair $(\mathcal{A}, \mathcal{D})$. Thus the support relation $\mathcal{S}$ is only used to constrain the defeat relation $\mathcal{D}$. [Prakken, 2014] calls the resulting frameworks *SuppAFs* and notes that $ASPIC^+$ can be reconstructed as an instance of *SuppAFs* as follows. Take $\mathcal{D}$ to be $ASPIC^+$'s defeat relation and $\mathcal{S}$ to be $ASPIC^+$'s subargument relation between arguments. It is then immediate from Definitions 2.10 and 2.12 that $ASPIC^+$'s notion of defeat satisfies [Dung and Thang, 2014]'s constraint on $\mathcal{D}$ in terms of $\mathcal{S}$.

An equivalent reformulation of *SuppAFs* does make use of support relations in its semantics. In [Prakken, 2013] $ASPIC^+$ as presented above was reformu-

[17] Where the requirement that an argument ordering is reasonable is adapted to the setting of EAFs.

lated in terms of [Pollock, 1994]'s recursive labellings, and this reformulation was abstracted to *SuppAFs* in [Prakken, 2014]. First, [Prakken, 2013] defines a notion of p-defeat (for "Pollock-defeat"), which captures direct defeat between arguments:

Definition 4.9 *[p-Attack] A* p-attacks *B iff A* p-undercuts*,* p-rebuts *or* p-undermines *B, where:*
- *A* p-undercuts *argument B iff $\texttt{Conc}(A) = -n(r)$ and B has a defeasible* top *rule r.*
- *A* p-rebuts *argument B iff $\texttt{Conc}(A) = -\texttt{Conc}(B)$ and B has a defeasible* top *rule.*
- *Argument A* p-undermines *B iff $\texttt{Conc}(A) = -\varphi$ and $B = \varphi$, $\varphi \notin \mathcal{K}_n$.*

Definition 4.10 *[p-Defeat] A* p-defeats *B iff:A p-undercuts B, or; A p-rebuts/p-undermines B and $A \nprec B$.*

Then [Prakken, 2013] proves that A defeats B according to Definition 2.12 iff A p-defeats B or A p-defeats a proper subargument B' of B. Now if the support relation of a SuppAF is taken to be *ASPIC*$^+$'s notion of an 'immediate' subargument and the defeat relation of a *SuppAF* is taken to be p-defeat, then the following definition is equivalent to [Dung, 1995]'s semantics for AFs (and so for *SuppAFs*).

Definition 4.11 *[p-labellings for* SuppAFs*.] Let $(\mathcal{A}, \mathcal{D}, \mathcal{S})$ be a* SuppAF *corresponding to a (c-)$SAF = (\mathcal{A}, \mathcal{D})$ where $\mathcal{D}$ is defined as p-defeat and where $\mathcal{S}$ is defined as $(A, B) \in \mathcal{S}$ iff B is of the form $B_1, \ldots, B_n \rightarrow / \Rightarrow \varphi$ and $A = B_i$ for some $1 \leq i \leq n$. Then (In, Out) is a* p-labelling *of* SuppAF *iff $In \cap Out = \emptyset$ and for all $A \in \mathcal{A}$ it holds that:*

1. *A is labelled in iff:*
   (a) *All arguments that p-defeat A are labelled out; and*
   (b) *All B that support A are labelled in.*
2. *A is labelled out iff:*
   (a) *A is p-defeated by an argument that is labelled in; or*
   (b) *Some B that supports A is labelled out.*

Exploiting the well-known correspondences between labelling- and extension-based semantics [Caminada, 2006], [Prakken, 2014] shows that the complete extensions defined thus for *SuppAFs* generated from *ASPIC*$^+$ with p-defeat are exactly the complete extensions of *SuppAFs* as generated above from *ASPIC*$^+$ with defeat.

[Prakken, 2014] also showed for preferred semantics that *ASPIC*$^+$ instantiates [Oren and Norman, 2008]'s evidential argumentation systems. One might expect that classical-logic instantiations of $ASPIC^+$ instantiate [Boella *et al.*, 2010a]'s version of bipolar argumentation frameworks for "deductive support". However, [Prakken, 2014] showed that this is not the case. This raises the question as to how one might instantiate [Boella *et al.*, 2010a]'s notion of deductive support.

More generally, the question arises as to the relation of the various accounts of abstract support relations with formalisms for structured argumentation. To the best of our knowledge, the only papers studying this question are [Prakken, 2014] and [Modgil, 2014]. [Modgil, 2014] discusses this issue under the assumption that arguments and their relations are constructed from a $ASPIC^+$ argumentation theory. He discusses how examples in the literature used to motivate the need for support relations essentially amount to the supporting argument A concluding some ϕ that is: 1) either a premise in the supported argument B; 2) the conclusion of a defeasible rule in B, or; 3) A provides the missing sub-argument for the *enthymeme* B (i.e., B is an incomplete argument). For example, letting A be an argument constructed from α and $\alpha \Rightarrow_{r_1} \beta$ then illustrating the three cases, B consists of: 1) β and $\beta \Rightarrow_{r_2} \delta$; 2) γ, $\gamma \Rightarrow_{r_3} \beta$ and $\beta \Rightarrow_{r_2} \delta$; 3) $\beta \Rightarrow_{r_2} \delta$.

Given this analysis, the underlying premises and rules can then be seen to generate additional arguments without the need for support relations; for example, in case 1) the additional argument $B' : A \Rightarrow_{r_2} \delta$. Hence, one would expect that the justification status of arguments obtained by the modified definitions of acceptability in abstract argumentation frameworks augmented by support relations, corresponds to their evaluation in a standard abstract argumentation framework of arguments and attacks, instantiated by the additional arguments generated by the same premises and rules. In case 1), this would mean that the status of B in the augmented framework in which B is supported by A, is the same as the status of B in the original framework consisting of A, B and B'. However, [Modgil, 2014] shows that this correspondence does not always hold[18]. He concludes from this that only when examining abstract concepts in a structured approach can one gain some insight into the appropriate use of these abstract level concepts in evaluating arguments. Indeed, [Modgil, 2014] provides a similar analysis of collective attacks [Nielsen and Parsons, 2007] and recursive attacks on attacks [Baroni *et al.*, 2011a] that have been incorporated at the abstract level and that have led to modified definitions of acceptability.

4.6.3 Preference- and value-based argumentation frameworks

[Amgoud and Cayrol, 1998] added to abstract argumentation frameworks (AFs) a preference relation on $\mathcal{A}$, resulting in *preference-based argumentation frameworks* ($PAFs$), which are triples of the form $\langle \mathcal{A}, attacks, \preceq \rangle$. An argument A

[18]Note that [Modgil, 2014] is careful to acknowledge that these observations apply to the case where arguments and their relations are generated by instantiating sets of formulae, rather than by human authoring of arguments. He argues that in the latter context additional relations between arguments incorporated in abstract argumentation frameworks may well be warranted by human oriented uses of argument, and goes on to argue the need for complementary empirical studies of human argumentation.

then *defeats* an argument B if A attacks B and $A \not\prec B$. Thus each PAF generates an AF of the form $\langle \mathcal{A}, \mathit{defeats} \rangle$, to which Dung's theory of AFs can be applied. [Bench-Capon, 2003] proposed a variant called *value-based argumentation frameworks* ($VAFs$), in which each argument is said to promote some (legal, moral or societal) value. Attacks in an $VAFs$ succeed only if the value promoted by the attacked argument is strictly preferred to the value of the attacking argument, according to a given ordering on the values (an *audience*).

The question arises as to what happens if $ASPIC^+$ is reformulated so as to generate $PAFs$ instead of Dung's original AFs. This can be easily done, since $ASPIC^+$ instantiations already generate a set of arguments with an attack relation and define a binary argument ordering. However, this may lead to violation of rationality postulates, even in cases where $ASPIC^+$ satisfies them.

Consider the following example from [Prakken, 2012b; Modgil and Prakken, 2013].

$A : p$
$B_1 : \neg p$
$B_2 : B_1 \Rightarrow q$

Here p and $\neg p$ are ordinary premises. Note that B_1 is a subargument of B_2. In $ASPIC^+$ we then have that A and B_1 directly attack each other while, moreover, A indirectly attacks B_2, since it directly attacks B_2's subargument B_1. These attack relations are displayed in Figure 5(a).

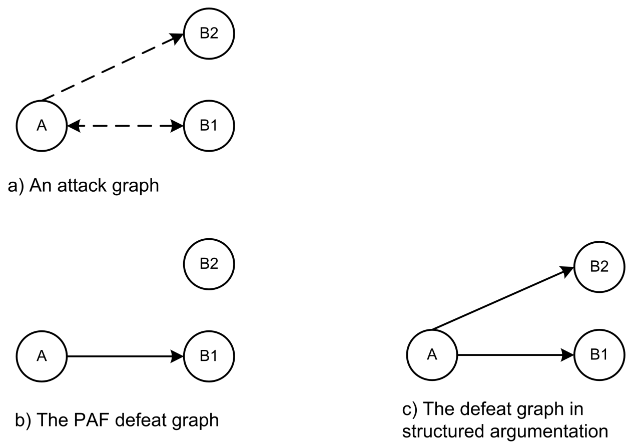

Figure 5. The attack graph

Assume next that $B_1 \prec A$ and $A \prec B_2$ (such an ordering could be the result of a last-link ordering). The PAF modelling then generates the following single defeat relation: A defeats B_1; see Figure 5(b). Then we have a single extension (in whatever semantics), namely, $\{A, B_2\}$. So not only A but also B_2 is justified.

However, this violates [Caminada and Amgoud, 2007]'s rationality postulate of subargument closure of extensions, since B_2 is in the extension while its subargument B_1 is not. This problem is not restricted to subargument closure; [Prakken, 2012b] also discusses examples in which the postulate of indirect consistency is violated.

The cause of the problem is that the PAF modelling of this example cannot recognise that the reason why A attacks B_2 is that A directly attacks B_1, which is a subargument of B_2. So the PAF modelling fails to capture that in order to check whether A's attack on B_2 succeeds, we should compare A not with B_2 but with B_1. Now since $B_1 \prec A$, then in $ASPIC^+$ we also have that A defeats B_2; see Figure 5(c). So the single extension (in whatever semantics) is $\{A\}$, and so closure under subarguments is respected.

This shows that under the assumption that $PAFs$ (and also $VAFs$) are instantiated by logical formulae, then these only behave correctly with respect to the rationality postulates, if all attacks are direct. We can conclude that for a principled analysis of the use of preferences to resolve attacks, the structure of arguments must be made explicit, since the structure of arguments is crucial in determining how preferences must be applied to attacks.

A more general word of caution is in order here. Although it is tempting to extend abstract argumentation frameworks with additional elements, one should resist the temptation to think that for any given argumentation phenomenon the most principled analysis is at the level of abstract argumentation. In fact, it often is the other way around, since at the abstract level crucial notions like claims, reasons and grounds are abstracted away.

4.6.4 Dynamics of argumentation

Recently much work has been done on the nature and effects of change operations on a given argumentation state, e.g. [Modgil, 2006; Baroni and Giacomin, 2008; Rotstein *et al.*, 2008; Baumann and Brewka, 2010; Cayrol *et al.*, 2010; Boella *et al.*, 2010b; Baroni *et al.*, 2011b]. Among other things, enforcing and preservation properties are studied. Enforcement concerns the extent to which desirable outcomes can or will be obtained by changing an argumentation state, while preservation is about the extent to which the current status of arguments is preserved under change. Almost all this work is done for abstract argumentation frameworks. In particular, the following operations on abstract argumentation frameworks have been studied: addition or deletion of (sets of) arguments and addition or deletion of (sets of) attack relations. Deleting attacks can here be seen as an abstraction from the use of preferences to resolve attacks into defeats.

The question arises as to what extent this work is relevant for $ASPIC^+$. Here too our above word of caution applies. At first sight, it would seem that the most principled analysis of argumentation dynamics is at the level of abstract argumentation frameworks. However, upon closer inspection it turns out that such analyses, because they ignore the structure of arguments, often implicitly make assumptions that are not in general satisfied by $ASPIC^+$

instantiations (and neither by other formalisms for structured argumentation). For example, abstract models of argumentation dynamics do not recognise that some arguments are not attackable (such as deductive arguments with certain premises) or that some attacks cannot be deleted (for example between arguments that were determined to be equally strong), or that the deletion of one argument implies the deletion of other arguments (when the deleted argument is a subargument of another, as in Figure 5 above), or that the deletion or addition of one attack implies the deletion or addition of other attacks (for example, attacking an argument implies that all arguments of which the attacked argument is a subargument are also attacked; in Figure 5 above attacking B_1 implies attacking B_2). These considerations imply that formal results pertaining to the abstract model are only relevant for specific cases, and fail to cover many realistic situations in argumentation that can be expressed in $ASPIC^+$. To give a very simple example, in models that allow the addition of arguments and attacks, any non-selfattacking argument A can be made a member of every extension by simply adding non-attacked attackers of all A's attackers. However, this result at the abstract level does not carry over to instantiations in which not all arguments are attackable. Here too, we see the importance of being aware of what the model abstracts from.

For these reasons we have in [Modgil and Prakken, 2012] proposed a model of preference dynamics in $ASPIC^+$, that arguably overcomes several limitations of [Baroni *et al.*, 2011b]'s resolution-based semantics for abstract argumentation frameworks when applied to preference-based dynamics.[19] The latter allows that symmetric attacks are replaced by asymmetric attacks (i.e., the symmetric attacks are 'resolved'). We argued that from the perspective of instantiated abstract argumentation frameworks, it is the use of preferences that provides the clearest motivation for obtaining resolutions. But then studying the use of preferences at the structured $ASPIC^+$ level suggests that one must also account for the resolution of asymmetric attacks, that preferences may also result in removal of both attacks in a symmetric attack, and that certain resolutions may be impossible, because assuming a preference that removes one attack may necessarily imply removal of another attack, or because some attacks cannot be removed by preferences (e.g. undercut attacks and attacks on contraries). These subtleties can only be appreciated at the structured level, and are thus not addressed by the study of resolutions at the abstract level adopted by [Baroni *et al.*, 2011b], in which only resolutions of symmetric attacks are considered, and all possible resolutions are considered possible.

5 Further Developments of ASPIC+

In Section 2 we presented what we called the 'basic' $ASPIC^+$ framework in two stages, first with symmetric negation and then generalising it with possibly asymmetric negation. As a matter of fact, this basic framework is the result

[19]We recognise that there may be other uses of resolution-based semantics to which our criticism does not apply.

of various revisions and incremental extensions [Amgoud *et al.*, 2006; Prakken, 2010; Modgil and Prakken, 2013; Modgil and Prakken, 2014]. Also, in [Modgil and Prakken, 2013], the basic framework in fact comes in four variants, resulting from whether the premises of arguments are assumed to be c-consistent or not and whether conflict-freeness is defined with the attack or the defeat relation (recall footnote 7). So instead of a single $ASPIC^+$ framework there in fact exists a family of such frameworks. And this family is growing. In this section we discuss recent work that modifies the $ASPIC^+$ framework in some respects, especially with new constraints on arguments or with modified or generalised notions of attack. We consider this development of variants of $ASPIC^+$ a healthy situation, since it amounts to a systematic investigation of the effects of different design choices within a common approach, which may each be applicable to certain kinds of problems.

5.1 Consistency and chaining restrictions motivated by contamination problems

Some recent work on $ASPIC^+$ has studied further constraints on arguments in an attempt to address the so-called contamination problem originally discussed by [Pollock, 1994; Pollock, 1995].[20] This problem arises if the strict inference rules are chosen to correspond to classical logic and if they are then combined with defeasible rules. The problem is how the trivialising effect of the classical Ex Falso principle can be avoided when two arguments that use defeasible rules have contradictory conclusions. The problem is especially hard since any solution should arguably preserve satisfaction of the rationality postulates of [Caminada and Amgoud, 2007]. In addition, [Caminada *et al.*, 2012] claim that any solution should also satisfy a new set of postulates that are meant to express the idea that information irrelevant to a part of the argumentation system should not affect the conclusions drawn from that part.

The following abstract example illustrates the problem. Assume that the strict rules of an argumentation system correspond to classical logic, i.e. $X \rightarrow \varphi \in \mathcal{R}_s$ if and only if $X \vdash \varphi$ and X is finite (where $\vdash$ denotes classical consequence).

Example 5.1 Let $\mathcal{R}_d = \{p \Rightarrow q;\ r \Rightarrow \neg q;\ t \Rightarrow s\}$, $\mathcal{K}_p = \emptyset$ and $\mathcal{K}_n = \{p, r, t\}$, while $\mathcal{R}_s$ corresponds to classical logic. Then the corresponding abstract argumentation framework includes the following arguments:

A_1: p $\quad A_2$: $A_1 \Rightarrow q$
B_1: r $\quad B_2$: $B_1 \Rightarrow \neg q$ $\quad C$: $A_2, B_2 \rightarrow \neg s$
D_1: t $\quad D_2$: $D_1 \Rightarrow s$

Figure 6 displays these arguments and their attack relations. Argument C attacks D_2. Whether C defeats D_2 depends on the argument ordering but

[20] Some parts of this section have been taken or adapted from [Grooters and Prakken, 2016].

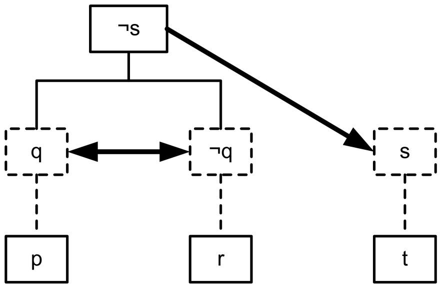

Figure 6. Illustrating trivialisation

plausible argument orderings are possible in which $C \not\prec D_2$ and so C defeats D_2. This is problematic, since s can be any formula, so any defeasible argument unrelated to A_2 or B_2, such as D_2, can, depending on the argument ordering, be defeated by C. Clearly, this is extremely harmful, since the existence of just a single case of mutual rebutting attack, which is very common, could trivialise the system. For instance, in this example neither of A_2 nor B_2 are in the grounded extension, since they defeat each other. But then the grounded extension does not defend D_2 against C and therefore does not contain D_2.

It should be noted that simply disallowing application of strict rules to inconsistent sets of formulas does not help, since then an argument for $\neg s$ can still be constructed as follows:

A_3: $A_2 \rightarrow q \vee \neg s$
C': $A_3, B_2 \rightarrow \neg s$

Note that argument C' does not apply any strict inference rule to an inconsistent set of formulas.

[Grooters and Prakken, 2016] propose the following formalisation of the property of trivialisation.

Definition 5.2 (Trivialising argumentation systems) *An argumentation system AS is* trivialising *iff for all $\varphi, \psi \in \mathcal{L}$ and all knowledge bases $\mathcal{K}$ such that $\{\varphi, \neg\varphi\} \subseteq \mathcal{K}$ a strict argument on the basis of $\mathcal{K}$ can be constructed in AS with conclusion ψ.*

The research problem then is identifying classes of non-trivialising argumentation systems. The argumentation system in our example is clearly trivialising since $\mathcal{R}_s$ contains strict rules $\varphi, \neg\varphi \rightarrow \psi$ for all $\varphi, \psi \in \mathcal{L}$.

Example 5.1 does not cause any problems for preferred or stable semantics, since A_2 and B_2 attack each other and at least one of these attacks will (with non-circular argument orderings) succeed as defeat. Therefore, all preferred or stable extensions contain either A_2 or B_2 but not both. Since both A_2 and B_2 attack C (by directly attacking one of its subarguments), C is for each preferred or stable extension defeated by at least one argument in the extension, so C is not in any of these extensions, so D_2 is in all these extensions. This is intuitively correct since there is no connection between D_2 and the arguments A_2 and B_2. [Pollock, 1994; Pollock, 1995] thought that this line of reasoning for preferred semantics suffices to show that his recursive-labelling approach (which was later in [Jakobovits and Vermeir, 1999] proved to be equivalent to preferred semantics) adequately deals with this problem. However, [Caminada, 2005] showed that the example can be extended in ways that also cause problems for preferred and stable semantics. Essentially, he replaced the facts p and r with defeasible arguments for p and r and let both these arguments be defeated by a self-defeating argument. On the one hand, such self-defeating arguments cannot be in any extension, since extensions are conflict free. However, if a self-defeating argument is not defeated by other arguments, it prevents any argument that it defeats from being acceptable with respect to an extension. In our example, if both A_2 and B_2 are defeated by a self-defeating argument that is otherwise undefeated, then neither A_2 not B_2 is in any extension, so no argument in an extension defends D_2 against C. To solve the problem, two approaches are possible. One is to change the definitions of the argumentation formalism, while the other is to derive the strict inference rules from a weaker logic than classical logic.

The first approach is taken by [Wu, 2012] and [Wu and Podlaszewski, 2015], who for the $ASPIC^+$ framework require that for each argument the set of conclusions of all its subarguments are classically consistent. They show that this solution partially works for a restricted version of $ASPIC^+$ without preferences, in that for complete semantics, both the original postulates of [Caminada and Amgoud, 2007] and the new ones of [Caminada *et al.*, 2012] are satisfied. However, their results do not cover stable, preferred or grounded semantics, while they give counterexamples to the consistency postulates for the case with preferences.

A second approach to solve the problem is to replace classical logic as the source for strict rules with a weaker, monotonic paraconsistent logic, in order to invalidate the Ex Falso principle as a valid strict inference rule. [Grooters and Prakken, 2016] explored this possibility. They first showed that two well-known paraconsistent logics, the system C_ω of [Da Costa, 1974] and the Logic of Paradox of [Priest, 1979; Priest, 1989], cannot be used for these purposes, since they induce violation of the postulate of indirect consistency. They then investigated Rescher and Manor's 1970 paraconsistent consequence notion of *weak consequence*. A set S of wff's weakly' implies a wff φ just in case at least one consistent subset of S classically implies φ. While thus initially taking the second approach, [Grooters and Prakken, 2016] had to combine it with the first approach (changing the definitions). Chaining strict rules in arguments has to be disallowed since the notion of weak consequence does not satisfy the Cut rule. For a counterexample, consider the set $\Gamma = \{a, \neg a \wedge b\}$. Then $\Gamma \vdash_W b$ and $\Gamma, b \vdash_W a \wedge b$, while it is not the case that $\Gamma \vdash_W a \wedge b$.

[Grooters and Prakken, 2016] proved that this solution avoids trivialisation and for well-behaved c-$SAFs$ satisfies all closure and consistency postulates (where the strict-closure postulate has to be changed to closure under one-step application of strict rules). Illustrating their solution with the above example, we see that the contaminating argument C cannot be constructed since its conclusion $\neg s$ follows from no consistent subset of $\{q, \neg q\}$, while the contaminating argument C' cannot be constructed since it chains two strict rules.

[Grooters and Prakken, 2016] also showed that with [Wu and Podlaszewski, 2015]'s stronger condition that the set of all conclusions of all subarguments of an argument must be consistent, consistency and strict closure are not satisfied. [Grooters and Prakken, 2016] did not attempt to prove Caminada *et al.*'s 2012 'contamination' postulates, for two reasons. First, they wanted to obtain results for all of [Dung, 1995]'s semantics and, second, they argued that Caminada *et al.*'s postulates in fact capture a stronger intuitive notion than the notion of trivialisation.

The work of [Grooters and Prakken, 2016] gives rise to some more general observations on [Caminada and Amgoud, 2007]'s original postulate of closure under strict rules. Above we suggested that $\mathcal{R}_s$ can be chosen to correspond to any monotonic logic with consequence notion $\vdash$ by letting $S \rightarrow \varphi \in \mathcal{R}_s$ if and only if $S \vdash \varphi$ and S is finite. However, the fact that the weak-consequence notion $\vdash_W$ does not satisfy the Cut rule illustrates that when $\mathcal{R}_s$ is thus defined, a system that is closed under $\mathcal{R}_s$ as defined in Section 3.1.2, could allow for inferences that are invalid according to $\vdash$. For these reasons, [Grooters and Prakken, 2016] not only reformulated their definition of strict closure but also proposed a new rationality postulate of *logical closure* and showed that their adapted version of $ASPIC^+$ also satisfies this postulate for well-behaved c-$SAFs$.

We also briefly note that [Grooters and Prakken, 2016] also studied minimality constraints on strict-rule applications and the exclusion of circular arguments. They show that if these two constraints are combined with their adoption of weak consequence as the source of the strict rules, then if both the knowledge base and the set of defeasible rules is finite, then each argument has at most a finite number of attackers, i.e., their framework generates so-called finitary argumentation frameworks in the sense of [Dung, 1995], which is computationally beneficial.

Finally, [D'Agostino and Modgil, 2016] provide a formalisation of classical argumentation with preferences in which arguments are triples (Δ, Γ, α) such that α is classically entailed by $\Delta \cup \Gamma$[21], and where Δ are the premises assumed true, and Γ the premises supposed true 'for the sake of argument'. The idea is that if a trivialising argument $(\{q, \neg q\}, \emptyset, s)$ defeats $(\{s\}, \emptyset, s) \in E$ (where E is an extension under any semantics), then $Y = (\emptyset, \{q, \neg q\}, \bot)$ defeats

[21] [D'Agostino and Modgil, 2016] allow for arguments with inconsistent premises, as they argue that arguments with inconsistent premises, and hence the trivialising effect of such arguments, should be excluded dialectically (as in real-world reasoning and debate), rather than checking for consistency prior to inclusion of the argument in an abstract argumentation framework.

$X = (\{q, \neg q\}, \emptyset, s)$ (Y supposes for the sake of argument the premises of X). Moreover, since the premises whose truth Y commits to are empty, Y cannot be defeated and so can be included in any E in order to defend $(\{s\}, \emptyset, s)$, thus negating the trivialising effect of X. [D'Agostino and Modgil, 2016] then show that under certain conditions, the consistency and closure postulates, as well as Caminada *et al.*'s additional contamination postulates are satisfied. As the authors note, an interesting direction for future research would be to see if their approach can be applied to the full $ASPIC^+$ framework.

5.2 Dung (2016) on rule-based argumentation systems

Recently, [Dung, 2016] has continued the formal study of [Dung and Thang, 2014]'s rule-based argumentation systems. Recall that these comprise of strict and defeasible inference rules over a propositional literal language, where axiom, respectively ordinary, premises p are simulated with rules $\rightarrow p$ and $\Rightarrow p$. [Dung, 2016] adds a transitive preference relation $\leq$ on $\mathcal{R}_d$, so that he defines rule-based systems as a triple $(\mathcal{R}_s, \mathcal{R}_d, \leq)$. In addition, he confines his study to knowledge bases with a consistent strict closure. Above we explained that [Dung and Thang, 2014] adopt the $ASPIC^+$ definitions of argument and defeat (which they call attack) and thus effectively study a class of $ASPIC^+$ instantiations. [Dung, 2016] also adopts the $ASPIC^+$ definition of an argument and still assumes that rule-based systems generate abstract argumentation frameworks in the sense of [Dung, 1995] (in our notation $(\mathcal{A}, \mathcal{D})$). However, Dung now abstracts from particular definitions of defeat ($\mathcal{D}$) and instead defines properties that defeat relations should have, thus effectively generalising $ASPIC^+$ on its notion of defeat. He then studies conditions under which defeat relations satisfy these properties.

Since this work is quite recent, we confine ourselves to a brief summary and discussion. In doing so, we will replace Dung's term 'attack' with 'defeat', in order to be consistent with the terminology in this chapter. This replacement is justified since in [Dung, 2016] it is the attack relation in terms of which arguments are evaluated, so it plays the role of $ASPIC^+$'s defeat relation.

Dung introduces two new rationality postulates. His postulate for *attack monotonicity* informally says that strengthening an argument cannot eliminate an attack of that argument on another. Let us illustrate this with Figure 2, interpreting the horizontal arrows as defeat relations. Then this postulate says, for instance, that if D_4's argument C_2 for v is replaced with a necessary premise v (or in [Dung, 2016]'s case a strict rule $\rightarrow v$) or with a strict and firm argument from u to v, then the new version of D_4 still defeats B_2. Next, Dung's postulate of *credulous cumulativity* informally means that changing a conclusion of an argument in some extension to a necessary fact cannot eliminate that extension.

Dung then identifies several sets of conditions under which one or both of these postulates and/or the original postulates of [Caminada and Amgoud, 2007] are satisfied. For the details of these very valuable results we refer the reader to his own publication. Dung then continues by investigating several definitions of defeat in terms of the preference relation $\leq$ on $\mathcal{R}_d$ on whether they satisfy these various postulates. Since he also assumes here that strict arguments cannot be defeated, this part of his study effectively concerns instantiations of $ASPIC^+$ as defined above in Section 2.2. Here Dung obtains both positive and negative results. For example, elitist orderings as defined in [Modgil and Prakken, 2013] are shown to satisfy attack monotonicity but not credulous cumulativity and indirect consistency, while democratic orderings as defined in [Modgil and Prakken, 2013] and Definition 3.4 above are shown to satisfy credulous cumulativity and indirect consistency but not attack monotonicity. As for Dung's results on consistency, these are a special case of [Modgil and Prakken, 2013]'s results for democratic orderings but they contain counterexamples to their results for the elitist orderings. However, these counterexamples do not apply to [Prakken, 2010]'s original way to define the elitist orderings, which has been incorporated in the above Definition 3.4, or to the erratum to [Modgil and Prakken, 2013] (which is available online at `https://nms.kcl.ac.uk/sanjay.modgil/AIJfinalErratum`).

The question arises as to whether Dung's two new postulates really are desirable in general. Our answer is positive for attack monotonicity but, following [Prakken and Vreeswijk, 2002, section 4.4], negative for credulous cumulativity. The point is that strengthening a defeasible conclusion to an indisputable fact may make arguments stronger than before, which can give them the power to defeat other arguments that they did not defeat before. This may in turn result in the loss of the extension from which the conclusion was promoted to an indisputable fact. We illustrate this with [Dung, 2016]'s own example. Informally: professors normally teach, administrators normally do not teach, deans are normally professors and all deans are administrators (so with transposition anyone who is not an administrator is not a dean). The question is whether some particular dean teaches. In rules:

$$
\begin{array}{lll}
\mathit{Dean} \Rightarrow_{d1} \mathit{Professor} & \mathit{Professor} \Rightarrow_{d2} \mathit{Teach} & \mathit{Administrator} \Rightarrow_{d3} \neg\, \mathit{Teach} \\
\mathit{Dean} \rightarrow \mathit{Administrator} & \neg\, \mathit{Administrator} \rightarrow \neg\, \mathit{Dean} & \rightarrow \mathit{Dean}
\end{array}
$$

Assume further that $d_1 < d_3 < d_2$. We have the following arguments on whether the dean teaches:

$$
\begin{array}{llll}
A_1: & \rightarrow \mathit{Dean} & B_1: & \rightarrow \mathit{Dean} \\
A_2: & A_1 \Rightarrow_{d1} \mathit{Professor} & B_2: & B_1 \rightarrow \mathit{Administrator} \\
A_3: & A_2 \Rightarrow_{d2} \mathit{Teach} & B_3: & B_2 \Rightarrow_{d3} \neg\, \mathit{Teach}
\end{array}
$$

(A_1 and B_1 are, of course, the same argument; B_3 is called A_3 by [Dung, 2016], while he does not explicitly name A_1/B_1 and B_2.) With the elitist or democratic weakest-link ordering as defined in Definition 3.4 above, argument B_3 strictly defeats A_3, so in all semantics a unique extension is obtained in which the dean is a professor but does not teach.

Suppose now the defeasibly justified conclusion *Professor* is added as a fact. This gives rise to a new argument:

C_1: $\rightarrow$ *Professor*
C_2: $C_1 \Rightarrow_{d2}$ *Teach*

Now the elitist ordering yields that C_2 strictly defeats B_3, so again in all semantics a unique extension is obtained but now it contains that the dean teaches. So we have lost the original extension, which illustrates violation of credulous cumulativity.

In our opinion, this outcome is the intuitive one, since by adding *Professor* as a fact, we have promoted its status from a defeasibly justified conclusion to an indisputable fact; as a consequence, argument A_3 can be strengthened by replacing its defeasible subargument A_2 with the strict-and-firm subargument C_1; no wonder then that the thus strengthened argument C_2 has, unlike its weaker version A_3, the power to defeat B_3.

Despite this minor criticism, we believe that Dung's latest investigations are a very valuable addition to the study of rule-based argumentation.

5.3 Variants of rebutting attack

Several papers have considered alternative definitions of rebutting attack in which an argument can under specific conditions also be rebutted on the conclusions of strict inferences.

5.3.1 Unrestricted rebuts

In $ASPIC^+$ as presented so far, arguments can only be rebutted on conclusions of defeasible-rule applications. [Caminada and Amgoud, 2007] call this *restricted rebut*. They also study *unrestricted rebut*, which allows rebuttals on the conclusion of a strict inference provided that at least one of the argument's subarguments is defeasible. Their replacement of restricted with unrestricted rebut leads to a variant of their simplified version of $ASPIC^+$ (which is in fact equivalent to [Dung and Thang, 2014]'s rule-based systems). They prove that for grounded semantics the rationality postulates are (under the usual conditions) satisfied but they provide a counterexample for stable and preferred semantics, presented above in Section 3.3 with a modification of Example 3.1.

[Caminada *et al.*, 2014] argue in favour of unrestricted rebut on the grounds that this would lead to more natural presentations of dialogues. They argue that when applying argumentation in dialogical settings, the notion of restricted rebuts sometimes forces agents to commit to statements they have insufficient reasons to believe. In abstract terms, suppose an agent Ag_1 submitting an argument A whose top rule is a strict rule $s_1 = \alpha_1, \ldots, \alpha_n \rightarrow \alpha$, where for $i = 1 \ldots n$, α_i is an ordinary premise in A or the head of a defeasible rule in A. Now suppose Ag_2 has an argument B that defeasibly concludes $\neg\alpha$. Since B does not rebut A on α, then to attack A requires that Ag_2 construct, for some $i = 1 \ldots n$, an argument B' that extends B *and the arguments concluding* α_j, $j \neq i$, with the transposition $s_1^i = \alpha_1, \ldots, \alpha_{i-1}, \neg\alpha, \alpha_{i+1}, \alpha_n \rightarrow \neg\alpha_i$. But then Ag_2 is forced to commit to her interlocutors' arguments concluding α_j, $j \neq i$, for which she has no reasons to believe.

[Caminada *et al.*, 2014] give the following concrete example.

John: *"Bob will attend conferences AAMAS and IJCAI this year, as he has papers accepted at both conferences."*
Mary: *"That won't be possible, as his budget of £1000 only allows for one foreign trip."*

Formally, this discussion could be modelled using an argumentation theory with $\mathcal{R}_d \supseteq \{\text{accA} \Rightarrow \text{attA};\ \text{accI} \Rightarrow \text{attI};\ \text{budget} \Rightarrow \neg(\text{attA} \wedge \text{attI})\}$ and $\mathcal{R}_s \supseteq \{\rightarrow \text{accA};\ \rightarrow \text{accI};\ \rightarrow \text{budget};\ \text{attA}, \text{attI} \rightarrow \text{attA} \wedge \text{attI}\}$.

A direct formalisation of the above arguments is then:

J_1:	$\rightarrow \text{accA}$	M_1:	$\rightarrow \text{budget}$
J_2:	$J_1 \Rightarrow \text{attA}$	M_2:	$M_1 \Rightarrow \neg(\text{attA} \wedge \text{attI})$
J_3:	$\rightarrow \text{accI}$		
J_4:	$J_3 \Rightarrow \text{attI}$		
J_5:	$J_3, J_4 \rightarrow \text{attA} \wedge \text{attI}$		

In *ASPIC*$^+$, Mary's argument does *not* attack John's argument, since the conclusion Mary wants to attack $(\text{attA} \wedge \text{attI})$ is the consequent of a strict rule. Mary can only attack John's argument by attacking the consequent of one of the defeasible rules, that is, by uttering one of the following two statements.

Mary$'$: *"Bob can't attend AAMAS because he will attend IJCAI, and his budget does not allow him to attend both."*
Mary$''$: *"Bob can't attend IJCAI because he will attend AAMAS, and his budget does not allow him to attend both."*

The associated formal counterarguments are as follows.[22]

M_1:	$\rightarrow \text{budget}$		
M_2:	$M_1 \Rightarrow \neg(\text{attA} \wedge \text{attI})$		
J_3:	$\rightarrow \text{accI}$	J_1:	$\rightarrow \text{accA}$
J_4:	$J_3 \Rightarrow \text{attI}$	J_2:	$J_1 \Rightarrow \text{attA}$
M_5':	$M_2, J_4 \rightarrow \neg\text{attA}$	M_5'':	$M_2, J_2 \rightarrow \neg\text{attI}$

According to [Caminada *et al.*, 2014] the problem with this is that Mary does not know which of the two conferences Bob will attend, but *ASPIC*$^+$ with restricted rebut forces her to assert that Bob will attend one or the other. They argue that from the perspective of commitment in dialogue [Walton and Krabbe, 1995], this is unnatural.

[Caminada *et al.*, 2014] then define a restricted version of basic *ASPIC*$^+$ as presented above in Section 2.2 – which they call *ASPIC*$^-$ – that substitutes strict rules with empty antecedents for axiom premises, and defeasible rules

[22] Assuming $\mathcal{R}_s$ ito be closed under transposition, the fact that $\mathcal{R}_s$ contains $\text{attA}, \text{attI} \rightarrow \text{attA} \wedge \text{attI}$ implies that $\mathcal{R}_s$ also contains $\neg(\text{attA} \wedge \text{attI}), \text{attI} \rightarrow \neg\text{attA}$ and $\text{attA}, \neg(\text{attA} \wedge \text{attI}) \rightarrow \neg\text{attI}$.

with empty antecedents for ordinary premises. Moreover, $ASPIC^-$ allows unrestricted rebuts on the conclusions of strict rules. They then show that under the assumption of a *total* ordering on the defeasible rules, and assuming either the `Elitist` or `Democratic` set comparisons used in defining weakest- or last-link preferences, all of [Caminada and Amgoud, 2007]'s rationality postulates are satisfied for well-behaved $SAFs$, but only for the grounded semantics. They have thus generalised [Caminada and Amgoud, 2007]'s results for some specific cases with preferences.

5.3.2 Weak rebuts and an alternative view on the rationality postulates

[Prakken, 2016] studies a weaker version of unrestricted rebut, motivated by the general observation that deductive inferences may weaken an argument. His argument is that when a deductive inference is made from the conclusions of at least two 'fallible' (defeasible or plausible) subarguments, the deductive inference can be said to aggregate the degrees of fallibility of the individual arguments to which it is applied. This in turn means that the deductive inference may be less preferred than either of these subarguments, so that a successful attack on the deductive inference does not necessarily imply a successful attack on one of its fallible subarguments. And this in turn means that there can be cases where it is rational to accept a set of arguments that is not strictly closed and that violate indirect consistency. Note that this line of reasoning does not apply to cases where a deductive inference is applied to at most one fallible subargument: then the amount of fallibility of the new argument is exactly the same as the amount of fallibility of the single fallible argument to which the deductive inference is applied. Accordingly, [Prakken, 2016] defines *weak rebut* as allowing rebuttals on the conclusion of a strict inference, provided that the strict inference is applied to at least two fallible subarguments. Moreover, he argues that there are cases where argument orderings cannot be required to satisfy all properties of a reasonable argument ordering as defined in Definition 3.16.

[Prakken, 2016] illustrates this with the lottery paradox, a well-known paradox from epistemology, first discussed by [Kyburg, 1961]. Imagine a fair lottery with one million tickets and just one prize. If the principle is accepted that it is rational to accept a proposition if its truth is highly probable, then for each ticket T_i it is rational to accept that T_i will not win while at the same time it is rational to accept that exactly one ticket will win. If we also accept that everything that deductively follows from a set of rationally acceptable propositions is rationally acceptable, then we have two rationally acceptable propositions that contradict each other: we can join all individual propositions $\neg T_i$ into a big conjunction $\neg T_1 \wedge \ldots \wedge \neg T_{1,000,000}$ with one million conjuncts, which contradicts the certain fact that exactly one ticket will win.

Many views on this paradox exist. [Prakken, 2016] wants to formalise the view that for each individual ticket it is rational to accept that it will not win while at the same time it is not rational to accept the conjunction of these acceptable beliefs. He considers the following modelling of the lottery paradox in $ASPIC^+$. Let $\mathcal{L}$ be a propositional language built from the set of atoms $\{T_i \mid 1 \leq i \leq 1,000,000\}$. Then let X denote a well-formed formula $X_1 \veebar \ldots \veebar X_{1,000,000}$ where $\veebar$ is exclusive or and where each X_i is of one of the following forms:

- If $i = 1$ then $X_i = T_1 \wedge \neg T_2 \wedge \ldots \wedge \neg T_n$
- If $i = n$ then $X_i = \neg T_1 \wedge \neg T_2 \wedge \ldots \wedge \neg T_{n-1} \wedge T_n$
- Otherwise $X_i = \neg T_1 \wedge \ldots \wedge \neg T_{i-1} \wedge T_i \wedge \neg T_{i+1} \wedge \ldots \wedge \neg T_n$

Next we choose $\mathcal{K}_p = \{\neg T_i \mid 1 \leq i \leq 1,000,000\}$, $\mathcal{K}_n = \{X\}$, $\mathcal{R}_s$ as consisting of all propositionally valid inferences from finite sets and $\mathcal{R}_d = \emptyset$.

The following arguments are relevant for any i such that $1 \leq i \leq 1,000,000$.

$$\neg T_i \qquad \text{and} \qquad \neg T_1, \ldots, \neg T_{i-1}, \neg T_{i+1}, \ldots, \neg T_{1,000,000}, X \to T_i \text{ (call it } A_i)$$

[Prakken, 2016] then equates rational acceptability with sceptical justification (see Definition 2.18 above). Making $\neg T_i$ sceptically justified for all i requires for all i that $A_i \prec \neg T_i$, to prevent A_i from defeating $\neg T_i$. Then we have a single extension in all semantics containing arguments for all conclusions $\neg T_i$ but not for their conjunction.

Note that adopting the above argument ordering requires that Condition (2) of Definition 3.16 of reasonable argument orderings is dropped, since it excludes such an argument ordering. On the other hand, Condition (1) of Definition 3.16 can be retained. In particular, Condition (1.iii) captures that applying a strict rule to the conclusion of a single argument A to obtain an argument A' does not change the 'preferedness' of A' compared to A. This is reasonable in general, since A and A' have exactly the same set of fallible elements (ordinary premises and/or defeasible inferences). [Prakken, 2016] calls argument orderings that satisfy Condition (1) of Definition 3.16 *weakly reasonable* argument orderings. Finally, he proposes weakened versions of the postulates of strict closure and indirect consistency, according to which these properties are only required to hold for subsets of extensions with at most one fallible argument. He then proves that if weak rebut is allowed in addition to restricted rebut and argument orderings are required to be weakly reasonable, then the original postulate of direct consistency plus the weakened postulates of strict closure and indirect consistency are satisfied if AT is closed under contraposition or transposition and $\texttt{Prem}(A) \cup \mathcal{K}_n$ is indirectly consistent.

[Prakken, 2016] concludes with some general observations on the relation between deduction and justification. He argues to have shown that preservation of truth (the definition of deductively valid arguments) does not imply preservation of rational acceptance, since truth and rational acceptance are different things. However, he also argues that deduction still plays an important role in argumentation. Deductive inference rules are still available as argument construction rules and if an argument with a strict top rule has no attackers or all its attackers are less preferred, then the argument may still be sceptically justified. The specifics of the adopted argument ordering are essential here. For instance, in the lottery paradox the argument ordering might allow that application of the conjunction rule to a small number of conclusions $\neg T_i$ is still sceptically justified.

5.4 Attacks from sets of arguments to arguments

[Baroni *et al.*, 2015] consider a variant of $ASPIC^+$ by adapting an idea originally proposed by [Vreeswijk, 1997] in the context of his 'abstract argumentation systems', which are a predecessor of $ASPIC^+$. In Vreeswijk's systems a counterargument is in fact a *set* of arguments: a set Σ of arguments is *incompatible* with an argument τ iff the conclusions of $\Sigma \cup \{\tau\}$ give rise to a strict argument for $\perp$. [Baroni *et al.*, 2015] adapt this idea to $ASPIC^+$, where the 'nodes' of the abstract argumentation frameworks generated by the modification are sets of arguments instead of individual arguments. They then prove satisfaction of [Caminada and Amgoud, 2007]'s rationality postulates under similar conditions as in [Modgil and Prakken, 2013].

[Baroni *et al.*, 2015]'s proposal is motivated by criticism of the $ASPIC^+$ treatment of generalised contrariness relations. However, we believe that they just criticise specific uses of this generalised contrariness relation and that the problems they discuss can be avoided by proper definitions of contrariness. Nevertheless, their ideas are very interesting and also apply to basic $ASPIC^+$ with ordinary negation. For example, it would be interesting to see if their variant of $ASPIC^+$ provides an alternative way to model the examples discussed by [Caminada *et al.*, 2014]. More generally, it would be interesting to see if their variant of $ASPIC^+$ can be reconstructed as generating AFs that allow attacks from sets of arguments to arguments as in e.g. [Bochman, 2003].

6 Implementations and applications

6.1 Implementations

Various implementations of instantiations of $ASPIC^+$are available online, all with domain-specific inference rules defined over literal-like languages, and with argument orderings based on rule preferences.

The original ASPIC inference engine The original inference engine from the ASPIC project (designed by Matthew South on the basis of a prototype of Gerard Vreeswijk) is available online at `http://aspic.cossac.org/`, with a demonstrator with example inputs available at `http://aspic.cossac.org/ArgumentationSystem/`. Rules can be formulated over a language with predicate-logic literals with ordinary negation. The implementation allows for choosing between restricted and unrestricted rebut. The implementation of restricted rebut deviates from its formal definition in that it also allows rebuttals between two arguments that both have a strict top rule. Arguments can be evaluated alternatively with a last- and a weakest-link argument ordering and with sceptical grounded or credulous preferred semantics.

Visser's Epistemic and Practical Reasoner Wietske Visser took the ASPIC deliverable ([Amgoud *et al.*, 2006]) as the basis for her Epistemic and Practical Reasoner (EPR), available at `http://www.wietskevisser.nl/research/epr/`. Rules can be formulated over a language of propositional literals with ordinary negation, optionally augmented with a 'desirable' modality for modelling practical reasoning. EPR implements argument games for sceptical grounded and credulous preferred semantics, as well as [Prakken, 2006]'s game for combined epistemic and practical reasoning. It also implements as an option [Prakken, 2005]'s mechanism for accrual of arguments.

ArgTech's TOAST Mark Snaith of ArgTech at the University of Dundee, Scotland, developed an implementation called TOAST ([Snaid and Reed, 2012]) based on [Prakken, 2010], available at `www.arg-tech.org/index.php/toast-an-aspic-implementation/`. Rules can be formulated over a language of propositional literals with ordinary negation plus optionally a user-specified contrariness relation. TOAST allows for argument evaluation with an elitist weakest- or last-link ordering and in grounded, preferred, stable and semi-stable semantics. Interestingly, TOAST can receive input specified in the AIF format, so that it can be connected to argumentation tools that can export to AIF ([Bex *et al.*, 2013a]). More on this will be said in the following subsection.

6.2 Logical specifications of the Argument Interchange Format

There is substantial interest in the development of argumentation support tools enabling the structuring of individual arguments and the dialogical exchange of argument in offline and online tools supporting human reasoning and debate (for example see `www.arg-tech.org`). A key aim is to then organise human authored arguments into abstract argumentation frameworks, so ensuring that the assessment of arguments is formally and rationally grounded and enabling 'mixed initiative' argumentation integrating both machine and human authored arguments [Modgil *et al.*, 2013]. These developments, as well as the burgeoning interest in logic-based models of argument, have motivated formulation of a standardised format – the *Argument Interchange Format* (*AIF*) [Chesñevar *et al.*, 2006] – for representation of human authored arguments and arguments constructed in logic.

The *AIF* is an ontology that broadly speaking distinguishes between information (propositions and sentences) and schemes which are general patterns of reasoning such as applications of inference rules, or conflict or preferences between information. Instances of these information and schemes classes constitute nodes that can be organised into *AIF* graphs representing argumentation knowledge. In [Bex *et al.*, 2013b], two-way translations are defined between *AIF* graphs and both $ASPIC^+$ and E-$ASPIC^+$ argumentation theories, and a number of information preserving properties are proved in both cases. The latter essentially prove that given certain assumptions on the given *AIF* graphs, the translation functions are identity-preserving (i.e. translating from the *AIF* graph to (E-)$ASPIC^+$ and back again yields the same graph as we started out with).

One can then translate *AIF* representations of human authored arguments and their interactions defined in the above-mentioned argumentation support tools, and translate these to instantiations of (E-)$ASPIC^+$ so enabling evaluation under Dung's semantics. This is explored in [Bex *et al.*, 2013b], in which arguments and their interactions authored in the *Rationale* tool [ter Berg *et al.*, 2009] are translated to the *AIF* and then to $ASPIC^+$ arguments, attacks and defeats. In this way, $ASPIC^+$ is placed in the wider spectrum of not just formal but also philosophical and linguistic approaches to argumentation.

6.3 Other applications of $ASPIC^+$

$ASPIC^+$ has been applied both in purely theoretical models and in implemented architectures.

6.3.1 Theoretical applications

Some theoretical applications of $ASPIC^+$ amount to the formulation of sets of argument schemes for specific forms of reasoning in $ASPIC^+$. [van der Weide *et al.*, 2011] and [van der Weide, 2011] use a combination of $ASPIC^+$ and [Wooldridge *et al.*, 2006]'s system for meta-argumentation for specifying argument schemes for reasoning about preferences in argumentation-based decision making. [Bench-Capon and Prakken, 2010] and [Bench-Capon *et al.*, 2011] formulate argument schemes for policy debates in E-$ASPIC^+$. [Prakken *et al.*, 2015] and [Bench-Capon *et al.*, 2013], inspired by earlier AI & Law work of e.g. [Ashley, 1990] and [Aleven, 2003], model factor-based legal reasoning with precedents in $ASPIC^+$, with argument schemes formalised as defeasible rules and auxiliary definitions concerning (sets of) factors, their origins, their relations and their preferences as first-order axioms. This allows the formalisation of arguments like the following:

> **Plaintiff** The current case and precedent *Bryce* share pro-plaintiff factors $\{f_1, f_2\}$ and pro-defendant factors $\{f_3\}$, the pro-plaintiff factors outweigh the pro-defendant factors since *Bryce* was decided for the plaintiff; therefore, the current case should be decided for me.
>
> **Defendant** But unlike the current case, Bryce also contained pro-plaintiff factor f_4, so it is relevantly different from the current case, so the outcome of *Bryce* does not control the current case.
>
> **Plaintiff** But the current case contains factor f_5 and both f_4 and f_5 are a special case of the more abstract factor f_6, so this difference between *Bryce* and the current case is not relevant.

Other theoretical applications of $ASPIC^+$ concern case studies. [Prakken, 2012a] modelled the legal and evidential reasoning in the American *Popov v.*

Hayashi case, an ownerships dispute between two baseball fans about a baseball hit in the 500th homerun of a famous American baseball player. [Prakken, 2015] modelled a legislative debate and an American labour law dispute as argumentation-based decision making involving goals, values and preferences.

Finally, some theoretical applications use $ASPIC^+$ as a component of a more general reasoning model. [Müller and Hunter, 2012] used a simple instantiation of $ASPIC^+$ with no knowledge base, only defeasible rules and no preferences as a reasoning component in a formal model of decision making. [Prakken *et al.*, 2013] applied $ASPIC^+$ in a dialogue model of collaborative IT security risk assessment. Finally, [Timmer *et al.*, 2017] used $ASPIC^+$ for generating explanations of forensic Bayesian networks.

6.3.2 Applications in implemented architectures

Some implemented architectures proposed in the literature have used implementations of $ASPIC^+$ as a component. [Kok, 2013] used $ASPIC^+$ as the agent reasoning mechanism in a testbed for inter-agent deliberation dialogue, meant for testing whether the use of argumentation is beneficial to the individual agents or to the group to which they belong. This testbed is available online at `https://bitbucket.org/erickok/baidd`. [Toniolo *et al.*, 2015] used $ASPIC^+$ as a reasoning component in their *CISpaces* sensemaking tool for intelligence analysis. [Yun and Croitoru, 2016] used the original ASPIC inference engine for reasoning with possibly inconsistent ontologies in ontology-based data access. Finally, [van Zee *et al.*, 2016] used the TOAST implementation of $ASPIC^+$ as a component of a framework for rationalising goal models using argument diagrams.

7 Open problems and avenues for future research

The study of abstract rule-based argumentation with both strict and defeasible rules has a long history, ultimately going back to the seminal work of [Pollock, 1987], passing through intermediate stages [Simari and Loui, 1992; Pollock, 1995; Vreeswijk, 1997; Prakken and Sartor, 1997; Garcia and Simari, 2004] and currently consolidated in the work on $ASPIC^+$. As this chapter has shown, the approach is a fruitful one, a mature metatheory is developing and there is a growing number of implementations and applications. Yet many open questions and avenues for future research remain. Here we list some of the (in our opinion) most important ones.

- The study of argument preference relations and their properties is relatively underdeveloped. More can be done here, for example, relating argument orderings to work in decision theory or to probability theory (see also the next point), or combining different preference criteria for different kinds of problems, such as for epistemic versus practical reasoning.
- A recent research trend in formal argumentation is the combination of argumentation-based inference with probability theory. This is not sur-

prising, since argumentation has from the early days been proposed as a model for reasoning under uncertainty. One question that arises here is how characterisations of the strength or relative preference of arguments relate to probability theory. Much recent work on probabilistic argumentation assigns probabilities to arguments in abstract argumentation frameworks, as in [Li *et al.*, 2012; Hunter and Thimm, 2014]. However, assigning probabilities to arguments is problematic, since in probability theory probabilities are assigned to the truth of statements or to outcomes of events, and an argument is neither a statement nor an event. What is required here is a precise specification of what the probability of an argument means in terms of its elements. How to do this in the context of abstract rule-based argumentation is still largely an open question. A preliminary answer is given by [Hunter, 2013] but only for the case of classical-logic argumentation.

- The contamination problems referred to in Section 5.1 remain to be solved for the fully general $ASPIC^+$framework. As briefly discussed at the end of Section 5.1, the work of [D'Agostino and Modgil, 2016] suggests directions for future development of the $ASPIC^+$framework such that one can establish conditions under which the additional rationality postulates of [Caminada *et al.*, 2012] are satisfied.

- In contrast to abstract argumentation, the study of computational aspects of rule-based argumentation and the various ways it can be instantiated is seriously underdeveloped. Much work can still be done on algorithms and complexity results for rule-based argumentation involving defeasible rules and preferences.

- While there is a growing body of work on the dynamics of abstract argumentation, the work of [Modgil and Prakken, 2012] in $ASPIC^+$ is to our knowledge still the only account of the dynamics of structured argumentation. Much remains to be done here.

- Another important research topic is implementation of more expressive instantiations than those existing today. It would, for example, be interesting to integrate state-of-the art propositional, first-order or modal-logic theorem provers in $ASPIC^+$ implementations.

- Finally, with an eye to practical applications it is important to conduct comparative case studies involving various formalisms, such as $ASPIC^+$, assumption-based argumentation, Carneades or [Brewka and Woltran, 2010]'s abstract dialectical frameworks. It would be especially interesting to study issues like naturalness and conciseness of representations.

Acknowledgments

We would like to express our thanks to those who have worked on the precursors of the $ASPIC^+$ framework (reviewed at the beginning of this chapter), as well as those who have provided valuable feedback (including Peter Young and Leon van der Torre), and those who have taken the time and effort to critically and constructively study $ASPIC^+$ and as a consequence have helped us improve $ASPIC^+$; notably, Phan Minh Dung in [Dung, 2016] and especially Sjur Dyrkolbotn (personal communication), whose constructive observations motivated us to write the erratum (`https://nms.kcl.ac.uk/sanjay.modgil/AIJfinalErratum`) to [Modgil and Prakken, 2013].

BIBLIOGRAPHY

[Aleven, 2003] V. Aleven. Using background knowledge in case-based legal reasoning: a computational model and an intelligent learning environment. *Artificial Intelligence*, 150:183–237, 2003.

[Amgoud and Besnard, 2013] L. Amgoud and Ph. Besnard. Logical limits of abstract argumentation frameworks. *Journal of Applied Non-classical Logics*, 23:229–267, 2013.

[Amgoud and Cayrol, 1998] L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 1–7, 1998.

[Amgoud and Cayrol, 2002] L. Amgoud and C. Cayrol. A model of reasoning based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–215, 2002.

[Amgoud *et al.*, 2006] L. Amgoud, L. Bodenstaff, M. Caminada, P. McBurney, S. Parsons, H. Prakken, J. van Veenen, and G.A.W. Vreeswijk. Final review and report on formal argumentation system. Deliverable D2.6, ASPIC IST-FP6-002307, 2006.

[Ashley, 1990] K.D. Ashley. *Modeling Legal Argument: Reasoning with Cases and Hypotheticals.* MIT Press, Cambridge, MA, 1990.

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

[Baroni *et al.*, 2011a] P. Baroni, F. Cerutti, M. Giacomin, and G. Guida. Afra: Argumentation framework with recursive attacks. *International Journal of Approximate Reasoning*, 52:19–37, 2011.

[Baroni *et al.*, 2011b] P. Baroni, P.E. Dunne, and M. Giacomin. On the resolution-based family of abstract argumentation semantics and its grounded instance. *Artificial Intelligence*, 175:791–813, 2011.

[Baroni *et al.*, 2015] P. Baroni, M. Giacomin, and B. Lao. Dealing with generic contrariness in structured argumentation. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence*, pages 2727–2733, 2015.

[Baumann and Brewka, 2010] R. Baumann and G. Brewka. Expanding argumentation frameworks: Enforcing and monotonicity results. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 75–86. IOS Press, Amsterdam etc, 2010.

[Bench-Capon and Prakken, 2010] T.J.M. Bench-Capon and H. Prakken. A lightweight formal model of two-phase democratic deliberation. In R.G.F. Winkels, editor, *Legal Knowledge and Information Systems. JURIX 2010: The Twenty-Third Annual Conference*, pages 27–36. IOS Press, Amsterdam etc., 2010.

[Bench-Capon *et al.*, 2011] T.J.M. Bench-Capon, H. Prakken, and W. Visser. Argument schemes for two-phase democratic deliberation. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Law*, pages 21–30, New York, 2011. ACM Press.

[Bench-Capon *et al.*, 2013] T.J.M. Bench-Capon, H. Prakken, A.Z. Wyner, and K. Atkinson. Argument schemes for reasoning with legal cases using values. In *Proceedings of* *the Fourteenth International Conference on Artificial Intelligence and Law*, pages 13–22, New York, 2013. ACM Press.

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

[Besnard and Hunter, 2001] Ph. Besnard and A. Hunter. A logic-based theory of deductive arguments. *Artificial Intelligence*, 128:203–235, 2001.

[Besnard and Hunter, 2008] Ph. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, Cambridge, MA, 2008.

[Bex *et al.*, 2013a] F.J. Bex, J. Lawrence, M. Snaith, and C. Reed. Implementing the argument web. *Communications of the ACM*, 56:66–73, 2013.

[Bex *et al.*, 2013b] F.J. Bex, S.J. Modgil, H. Prakken, and C.A. Reed. On logical specifications of the argument interchange format. *Journal of Logic and Computation*, 23:951–989, 2013.

[Bochman, 2003] A. Bochman. Collective argumentation and disjunctive logic programming. *Journal of Logic and Computation*, 13:405–428, 2003.

[Boella *et al.*, 2010a] G. Boella, D.M. Gabbay, L. van der Torre, and S. Villata. Support in abstract argumentation. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 111–122. IOS Press, Amsterdam etc, 2010.

[Boella *et al.*, 2010b] G. Boella, S. Kaci, and L. van der Torre. Dynamics in argumentation with single extensions: attack refinement and the grounded extension (extended version). In P. McBurney, I. Rahwan, S. Parsons, and N. Maudet, editors, *Argumentation in Multi-Agent Systems, 6th International Workshop, ArgMAS 2009, Budapest, Hungary, May 12, 2009. Revised Selected and Invited Papers*, number 6057 in Springer Lecture Notes in AI, pages 150–159. Springer Verlag, Berlin, 2010.

[Bondarenko *et al.*, 1993] A. Bondarenko, R.A. Kowalski, and F. Toni. An assumption-based framework for non-monotonic reasoning. In *Proceedings of the second International Workshop on Logic Programming and Nonmonotonic Logic*, pages 171–189, Lisbon (Portugal), 1993.

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

[Brewka and Gordon, 2010] G. Brewka and T.F. Gordon. Carneades and abstract dialectical frameworks: A reconstruction. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 3–12. IOS Press, Amsterdam etc, 2010.

[Brewka and Woltran, 2010] G. Brewka and S. Woltran. Abstract dialectical frameworks. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference*, pages 102–111. AAAI Press, 2010.

[Brewka, 1989] G. Brewka. Preferred subtheories: An extended logical framework for default reasoning. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1043–1048, 1989.

[Brewka, 1994a] G. Brewka. Adding priorities and specificity to default logic. In C. MacNish, D. Pearce, and L. Moniz Pereira, editors, *Logics in Artificial Intelligence. Proceedings of JELIA 1994*, number 838 in Springer Lecture Notes in AI, pages 247–260, Berlin, 1994. Springer Verlag.

[Brewka, 1994b] G. Brewka. Reasoning about priorities in default logic. In *Proceedings of the 12th National Conference on Artificial Intelligence (AAAI-94)*, pages 247–260, 1994.

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

[Caminada *et al.*, 2012] M. Caminada, W.A. Carnielli, and P.E. Dunne. Semi-stable semantics. *Journal of Logic and Computation*, 22:1207–1254, 2012.

[Caminada *et al.*, 2014] M. Caminada, S. Modgil, and N. Oren. Preferences and unrestricted rebut. In S. Parsons, N. Oren, C. Reed, and F. Cerutti, editors, *Computational Models of Argument. Proceedings of COMMA 2014*, pages 209–220. IOS Press, Amsterdam etc, 2014.

[Caminada, 2004] M. Caminada. *For the sake of the Argument. Explorations into argument-based reasoning.* Doctoral dissertation Free University Amsterdam, 2004.

[Caminada, 2005] M. Caminada. Contamination in formal argumentation systems. In *Proceedings of the Seventeenth Belgian-Dutch Conference on Artificial Intelligence (BNAIC-05)*, Brussels, Belgium, 2005.

[Caminada, 2006] M. Caminada. On the issue of reinstatement in argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence. Proceedings of JELIA 2006*, number 4160 in Springer Lecture Notes in AI, pages 111–123, Berlin, 2006. Springer Verlag.

[Cayrol and Lagasquie-Schiex, 2005] C. Cayrol and M.-C. Lagasquie-Schiex. On the acceptability of arguments in bipolar argumentation. In L. Godo, editor, *Proceedings of the 8nd European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 05)*, number 3571 in Springer Lecture Notes in AI, pages 378–389, Berlin, 2005. Springer Verlag.

[Cayrol and Lagasquie-Schiex, 2009] C. Cayrol and M.-C. Lagasquie-Schiex. Bipolar abstract argumentation systems. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 65–84. Springer, Berlin, 2009.

[Cayrol and Lagasquie-Schiex, 2013] C. Cayrol and M.-C. Lagasquie-Schiex. Bipolarity in argumentation graphs: Towards a better understanding. *International Journal of Approximate Reasoning*, 54:876–899, 2013.

[Cayrol *et al.*, 2010] C. Cayrol, F Dupin de Saint-Cyr, and M.-C. Lagasquie-Schiex. Change in abstract argumentation frameworks: adding an argument. *Journal of Artificial Intelligence Research*, 38:49–84, 2010.

[Cayrol, 1995] C. Cayrol. On the relation between argumentation and non-monotonic coherence-based entailment. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 1443–1448, 1995.

[Chesñevar *et al.*, 2006] C.I. Chesñevar, J. McGinnis, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, and S. Willmott. Towards an argument interchange format. *The Knowledge Engineering Review*, 21:293–316, 2006.

[Da Costa, 1974] N.C.A. Da Costa. On the theory of inconsistent formal systems. *Notre Dame Journal of Formal Logic*, 15(4):497–510, 1974.

[D'Agostino and Modgil, 2016] M. D'Agostino and S. Modgil. A rational account of classical logic argumentation for real-world agents. In *Proceedings of the Twenty Second European Conference on Artificial Intelligence (ECAI 2016)*, pages 141–149, 2016.

[Delgrande and Schaub, 2000] J. Delgrande and T. Schaub. Expressing preferences in default logic. *Artificial Intelligence*, 123:41–87, 2000.

[Dung and Thang, 2014] P.M. Dung and P.M. Thang. Closure and consistency in logic-associated argumentation. *Journal of Artificial Intelligence Research*, 49:79–109, 2014.

[Dung *et al.*, 2007] P.M. Dung, P. Mancarella, and F. Toni. Computing ideal sceptical argumentation. *Artificial Intelligence*, 171:642–674, 2007.

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

[Dung, 2014] P.M. Dung. An axiomatic analysis of structured argumentation for prioritized default reasoning. In *Proceedings of the 21st European Conference on Artificial Intelligence*, pages 267–272, 2014.

[Dung, 2016] P.M. Dung. An axiomatic analysis of structured argumentation with priorities. *Artificial Intelligence*, 231:107–150, 2016.

[Garcia and Simari, 2004] A.J. Garcia and G.R. Simari. Defeasible logic programming: An argumentative approach. *Theory and Practice of Logic Programming*, 4:95–138, 2004.

[Gordon and Walton, 2009a] T.F. Gordon and D.N. Walton. Legal reasoning with argumentation schemes. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Law*, pages 137–146, New York, 2009. ACM Press.

[Gordon and Walton, 2009b] T.F. Gordon and D.N. Walton. Proof burdens and standards. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 239–258. Springer, Berlin, 2009.

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

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

[Grooters and Prakken, 2016] D. Grooters and H. Prakken. Two aspects of relevance in structured argumentation: minimality and paraconsistency. *Journal of Artificial Intelligence Research*, 56:197–245, 2016.

[Grooters, 2014] D. Grooters. Paraconsistent logics in argumentation systems. Master's thesis, Department of Information and Computing Sciences, Utrecht University, Utrecht, 2014.

[Hunter and Thimm, 2014] A. Hunter and M. Thimm. Probabilistic argumentation with incomplete information. In *Proceedings of the 21st European Conference on Artificial Intelligence*, pages 1033–1034, 2014.

[Hunter, 2010] A. Hunter. Base logics in argumentation. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 275–286. IOS Press, Amsterdam etc, 2010.

[Hunter, 2013] A. Hunter. A probabilistic approach to modelling uncertain logical arguments. *International Journal of Approximate Reasoning*, 54:47–81, 2013.

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

[Kok, 2013] E.M. Kok. *Exploring the Practical Benefits of Argumentation in Multi-Agent Deliberation.* Doctoral dissertation Department of Information and Computing Sciences, Utrecht University, 2013.

[Kowalski and Toni, 1996] R.A. Kowalski and F. Toni. Abstract argumentation. *Artificial Intelligence and Law*, 4:275–296, 1996.

[Kraus *et al.*, 1990] S. Kraus, D.J. Lehmann, and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44:167–207, 1990.

[Kyburg, 1961] H. Kyburg. *Probability and the Logic of Rational Belief.* Wesleyan U. P. Middletown, CT, 1961.

[Li and Parsons, 2015] Z. Li and S. Parsons. On argumentation with purely defeasible rules. In C. Beierle and A. Dekhtyar, editors, *Proceedings of the 9th International Conference on Scalable Uncertainty Management (SUM'15)*, number 9310 in Springer Lecture Notes in AI, pages 30–43, Berlin, 2015. Springer Verlag.

[Li *et al.*, 2012] H. Li, N. Oren, and T. Norman. Probabilistic argumentation frameworks. In S. Modgil, N. Oren, and F. Toni, editors, *Theorie and Applications of Formal Argumentation. First International Workshop, TAFA 2011. Barcelona, Spain, July 16-17, 2011, Revised Selected Papers*, number 7132 in Springer Lecture Notes in AI, pages 1–16, Berlin, 2012. Springer Verlag.

[Lin and Shoham, 1989] F. Lin and Y. Shoham. Argument systems. A uniform basis for nonmonotonic reasoning. In *Principles of Knowledge Representation and Reasoning: Proceedings of the First International Conference*, pages 245–255, San Mateo, CA, 1989. Morgan Kaufmann Publishers.

[Martinez *et al.*, 2006] D.C. Martinez, A.J. Garcia, and G.R. Simari. On acceptability in abstract argumentation frameworks with an extended defeat relation. In P.E. Dunne and T.B.C. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 273–278. IOS Press, Amsterdam etc, 2006.

[Modgil and Caminada, 2009] S. Modgil and M. Caminada. Proof theories and algorithms for abstract argumentation frameworks. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 105–129. Springer, Berlin, 2009.

[Modgil and Prakken, 2010] S. Modgil and H. Prakken. Reasoning about preferences in structured extended argumentation frameworks. In P. Baroni, F. Cerutti, M. Giacomin, and G.R. Simari, editors, *Computational Models of Argument. Proceedings of COMMA 2010*, pages 347–358. IOS Press, Amsterdam etc, 2010.

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

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

[Modgil and Prakken, 2014] S. Modgil and H. Prakken. The ASPIC+ framework for structured argumentation: a tutorial. *Argument and Computation*, 5:31–62, 2014.

[Modgil *et al.*, 2013] S. Modgil, F. Toni, F. Bex, I. Bratko, C. Chesñevar, W. Dvorak, M. Falappa, X. Fan, S. Gaggl, A. Garcia, M. Gonzalez, T. Gordon, J. Leite, M. Mozina, C. Reed, G. Simari, S. Szeider, P. Torroni, and S. Woltran. The added value of argumentation. In S. Ossowski, editor, *Agreement Technologies*, pages 357–403. Springer, 2013.

[Modgil, 2006] S. Modgil. Hierarchical argumentation. In M. Fischer, W. van der Hoek, B. Konev, and A. Lisitsa, editors, *Logics in Artificial Intelligence. Proceedings of JELIA 2006*, number 4160 in Springer Lecture Notes in AI, pages 319–332, Berlin, 2006. Springer Verlag.

[Modgil, 2009] S. Modgil. Reasoning about preferences in argumentation frameworks. *Artificial Intelligence*, 173:901–934, 2009.

[Modgil, 2014] S. Modgil. Revisiting abstract argumentation. In E. Black, S. Modgil, and N. Oren, editors, *Second International Workshop, TAFA 2013, Beijing, China, August 3-5, 2013, Revised Selected papers*, number 8306 in Springer Lecture Notes in AI, pages 1–15, Berlin, 2014. Springer Verlag.

[Müller and Hunter, 2012] J. Müller and A. Hunter. An argumentation-based approach for decision making. In *Proceedings of the 25th International Conference on Tools with Artificial Intelligence (ICTAI 2013)*, pages 564–571, 2012.

[Nielsen and Parsons, 2007] S.H. Nielsen and S. Parsons. A generalization of Dung's abstract framework for argumentation: Arguing with sets of attacking arguments. In S. Parsons, N. Maudet, and I. Rahwan, editors, *Argumentation in Multi-Agent Systems. Third International Workshop, ArgMAS 2006, Hakodate, Japan, May 8, 2006, Revised Selected and Invited Papers*, number 4766 in Springer Lecture Notes in AI, pages 54–73. Springer Verlag, Berlin, 2007.

[Nouioua and Risch, 2011] F. Nouioua and V. Risch. Argumentation frameworks with necessities. In *Proceedings of the 4th International Conference on Scalable Uncertainty Management (SUM'11)*, number 6929 in Springer Lecture Notes in AI, pages 163–176, Berlin, 2011. Springer Verlag.

[Oren and Norman, 2008] N. Oren and T.J. Norman. Semantics for evidence-based argumentation. In Ph. Besnard, S. Doutre, and A. Hunter, editors, *Computational Models of Argument. Proceedings of COMMA 2008*, pages 276–284, Amsterdam etc, 2008. IOS Press.

[Pearl, 1992] J. Pearl. Epsilon-semantics. In S.C. Shapiro, editor, *Encyclopedia of Artificial Intelligence*, pages 468–475. John Wiley & Sons, New York, 1992.

[Pollock, 1987] J.L. Pollock. Defeasible reasoning. *Cognitive Science*, 11:481–518, 1987.

[Pollock, 1990] J.L. Pollock. A theory of defeasible reasoning. *International Journal of Intelligent Systems*, 6:33–54, 1990.

[Pollock, 1992] J.L. Pollock. How to reason defeasibly. *Artificial Intelligence*, 57:1–42, 1992.

[Pollock, 1994] J.L. Pollock. Justification and defeat. *Artificial Intelligence*, 67:377–408, 1994.

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

[Pollock, 2009] J.L. Pollock. A recursive semantics for defeasible reasoning. In I. Rahwan and G.R. Simari, editors, *Argumentation in Artificial Intelligence*, pages 173–197. Springer, Berlin, 2009.

[Prakken and Sartor, 1997] H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics*, 7:25–75, 1997.

[Prakken and Vreeswijk, 2002] H. Prakken and G.A.W. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Günthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 219–318. Kluwer Academic Publishers, Dordrecht/Boston/London, second edition, 2002.

[Prakken *et al.*, 2013] H. Prakken, D. Ionita, and R. Wieringa. Risk assessment as an argumentation game. In J. Leite, T.C. Son, P. Torroni, L. van der Torre, and S. Woltran, editors, *Proceedings of the 14th International Workshop on Computational Logic in Multi-Agent Systems (CLIMA XIV)*, number 8143 in Springer Lecture Notes in AI, pages 357–373, Berlin, 2013. Springer Verlag.

[Prakken *et al.*, 2015] H. Prakken, A.Z. Wyner, T.J.M. Bench-Capon, and K. Atkinson. A formalisation of argumentation schemes for legal case-based reasoning in ASPIC+. *Journal of Logic and Computation*, 25:1141–1166, 2015.

[Prakken, 1999] H. Prakken. Dialectical proof theory for defeasible argumentation with defeasible priorities (preliminary report). In J.-J.Ch. Meyer and P.-Y. Schobbens, editors, *Formal Models of Agents*, number 1760 in Springer Lecture Notes in AI, pages 202–215, Berlin, 1999. Springer Verlag.

[Prakken, 2005] H. Prakken. A study of accrual of arguments, with applications to evidential reasoning. In *Proceedings of the Tenth International Conference on Artificial Intelligence and Law*, pages 85–94, New York, 2005. ACM Press.

[Prakken, 2006] H. Prakken. Combining sceptical epistemic reasoning with credulous practical reasoning. In P.E. Dunne and T.B.C. Bench-Capon, editors, *Computational Models of Argument. Proceedings of COMMA 2006*, pages 311–322. IOS Press, Amsterdam etc, 2006.

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

[Prakken, 2012a] H. Prakken. Reconstructing Popov v. Hayashi in a framework for argumentation with structured arguments and Dungean semantics. *Artificial Intelligence and Law*, 20:57–82, 2012.

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

[Prakken, 2013] H. Prakken. Relating ways to instantiate abstract argumentation frameworks. In K.D. Atkinson, H. Prakken, and A.Z. Wyner, editors, *From Knowledge Representation to Argumentation in AI, Law and Policy Making. A Festschrift in Honour of Trevor Bench-Capon on the Occasion of his 60th Birthday*, pages 167–189. College Publications, London, 2013.

[Prakken, 2014] H. Prakken. On support relations in abstract argumentation as abstractions of inferential relations. In *Proceedings of the 21st European Conference on Artificial Intelligence*, pages 735–740, 2014.

[Prakken, 2015] H. Prakken. Formalising debates about law-making proposals as practical reasoning. In M. Araszkiewicz and K. Płeszka, editors, *Logic in the Theory and Practice of Lawmaking*, pages 301–321. Springer, Berlin, 2015.

[Prakken, 2016] H. Prakken. Rethinking the rationality postulates for argumentation-based inference. In P. Baroni, T.F. Gordon, T. Scheffler, and M. Stede, editors, *Computational Models of Argument. Proceedings of COMMA 2016*, pages 419–430. IOS Press, Amsterdam etc, 2016.

[Priest, 1979] G. Priest. The logic of paradox. *Journal of Philosophical Logic*, 8(1):219–241, 1979.

[Priest, 1989] G. Priest. Reasoning about truth. *Artificial Intelligence*, 39(2):231–244, 1989.

[Reiter, 1980] R. Reiter. A logic for default reasoning. *Artificial Intelligence*, 13:81–132, 1980.

[Rescher and Manor, 1970] N. Rescher and R. Manor. On inference from inconsistent premises. *Journal of Theory and Decision*, 1:179–219, 1970.

[Rotstein *et al.*, 2008] N.D. Rotstein, M.O. Moguillansky, M.A. Falappa, A.J. Garcia, and G.R. Simari. Argument theory change: Revision upon warrant. In Ph. Besnard, S. Doutre, and A. Hunter, editors, *Computational Models of Argument. Proceedings of COMMA 2008*, pages 336–347, Amsterdam etc, 2008. IOS Press.

[Simari and Loui, 1992] G.R. Simari and R.P. Loui. A mathematical treatment of defeasible argumentation and its implementation. *Artificial Intelligence*, 53:125–157, 1992.

[Snaid and Reed, 2012] M. Snaid and C. Reed. TOAST: Online ASPIC+ argumentation. In B. Verheij, S. Woltran, and S. Szeider, editors, *Computational Models of Argument. Proceedings of COMMA 2012*, pages 509–510. IOS Press, Amsterdam etc, 2012.

[ter Berg *et al.*, 2009] T. ter Berg, T. van Gelder, F. Patterson, and S. Teppema. *Critical Thinking: Reasoning and Communicating with Rationale*. Pearson Education Benelux, Amsterdam, 2009.

[Timmer *et al.*, 2017] S. Timmer, J.-J.Ch. Meyer, H. Prakken, S. Renooij, and B. Verheij. A two-phase method for extracting explanatory arguments from Bayesian networks. *International Journal of Approximate Reasoning*, 80:475–494, 2017.

[Toni, 2014] F. Toni. A tutorial on assumption-based argumentation. *Argument and Computation*, 5:89–117, 2014.

[Toniolo *et al.*, 2015] A. Toniolo, T.J. Norman, A. Etuk, F. Cerutti, R. Wentao Ouyang, M. Srivastava, N. Oren, T. Dropps, J.A. Allen, and Paul Sullivan. Supporting reasoning with different types of evidence in intelligence analysis. In *Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'15)*, pages 781–789, 2015.

[van der Weide *et al.*, 2011] T. van der Weide, F. Dignum, J.-J.Ch. Meyer, H. Prakken, and G. Vreeswijk. Arguing about preferences and decisions. In P. McBurney, I. Rahwan, and S. Parsons, editors, *Argumentation in Multi-Agent Systems, 7th International Workshop, ArgMAS 2010, Toronto, Canada, May 2010. Revised Selected and Invited Papers*, number 6614 in Springer Lecture Notes in AI, pages 68–85. Springer Verlag, Berlin, 2011.

[van der Weide, 2011] T.L. van der Weide. *Arguing to Motivate Decisions.* Doctoral dissertation Department of Information and Computing Sciences, Utrecht University, 2011.

[Van Gijzel and Prakken, 2011] B. Van Gijzel and H. Prakken. Relating Carneades with abstract argumentation. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 1113–1119, 2011.

[Van Gijzel and Prakken, 2012] B. Van Gijzel and H. Prakken. Relating Carneades with abstract argumentation via the ASPIC+ framework for structured argumentation. *Argument and Computation*, 3:21–47, 2012.

[van Zee *et al.*, 2016] M. van Zee, D. Marosin, F.J. Bex, and S. Ghanavati. RationalGRL: A framework for rationalizing goal models using argument diagrams. In *Proceedings of the 35th International Conference on Conceptual Modeling (ER'2016)*, 2016.

[Verheij, 2003] B. Verheij. DefLog: on the logical interpretation of prima facie justified assumptions. *Journal of Logic and Computation*, 13:319–346, 2003.

[Vreeswijk, 1997] G.A.W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

[Walton and Krabbe, 1995] D.N. Walton and E.C.W. Krabbe. *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning.* State University of New York Press, Albany, NY, 1995.

[Walton, 1996] D.N. Walton. *Argumentation Schemes for Presumptive Reasoning.* Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[Walton, 2006] D.N. Walton. *Fundamentals of Critical Argumentation.* Cambridge University Press, Cambridge, 2006.

[Wooldridge *et al.*, 2006] M. Wooldridge, P. McBurney, and S. Parsons. On the meta-logic of arguments. In S. Parsons, N. Maudet, P. Moraitis, and I. Rahwan, editors, *Argumentation in Multi-Agent Systems: Second International Workshop, ArgMAS 2005, Utrecht, Netherlands, July 26, 2005, Revised Selected and Invited Papers*, number 4049 in Springer Lecture Notes in AI, pages 42–56. Springer Verlag, Berlin, 2006.

[Wu and Podlaszewski, 2015] Y. Wu and M. Podlaszewski. Implementing crash-resistence and non-interference in logic-based argumentation. *Journal of Logic and Computation*, 25:303–333, 2015.

[Wu, 2012] Y. Wu. *Between Argument and Conclusion. Argument-based Approaches to Discussion, Inference and Uncertainty.* Doctoral Dissertation Faculty of Sciences, Technology and Communication, University of Luxemburg, 2012.

[Young *et al.*, 2016] A.P. Young, S. Modgil, and O. Rodrigues. Prioritised default logic as rational argumentation. In *Proceedings of the 15th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'16)*, pages 626–634, 2016.

[Yun and Croitoru, 2016] B. Yun and M. Croitoru. An argumentation workflow for reasoning in ontology based data access. In P. Baroni, T.F. Gordon, T. Scheffler, and M. Stede, editors, *Computational Models of Argument. Proceedings of COMMA 2016*, pages 61–68. IOS Press, Amsterdam etc, 2016.

Sanjay Modgil
Department of Informatics, King's College London, UK
Email: sanjay.modgil@kcl.ac.uk

Henry Prakken
Department of Information and Computing Sciences, Utrecht University, The Netherlands
Faculty of Law, University of Groningen, The Netherlands
Email: H.Prakken@uu.nl