Clarifying some misconceptions on the *ASPIC*$^+$ framework

Henry Prakken [a] Sanjay Modgil [b]
[a] *Department of Information and Computing Sciences, University of Utrecht and Faculty of Law, University of Groningen, The Netherlands*
[b] *Department of Computer Science, Kings College London, UK*

Abstract. The *ASPIC*$^+$ framework is a general framework for argumentation-based inference which aims to unifies two research strands: those in which arguments can only be attacked on their defeasible inferences and those in which arguments can only be attacked on their premises. The framework is meant to define a wide class of instantiations of abstract argumentation frameworks and to support the investigation of rationality postulates for argumentation-based inference.

Recently, it has been argued that the *ASPIC*$^+$ framework suffers from several weaknesses. In this paper these criticisms are argued to be based on a number of misconceptions on the nature of the *ASPIC*$^+$ framework.

Keywords. Argumentation frameworks, *ASPIC*$^+$, Rationality postulates

1. Introduction

In [19,15] a general framework called *ASPIC*$^+$ is proposed for argumentation-based inference with two kinds of inference rules, strict and defeasible ones. The framework is meant to define a wide class of instantiations of [10]'s abstract argumentation frameworks, and to integrate and further develop two main research strands in the study of argument-based inference. The first research strand is work that locates all fallibility of arguments in the use of defeasible inference rules, such as John Pollock's work [18], Vreeswijk's abstract argumentation systems [22] and Defeasible Logic Programming [11]. The second approach is work that locates all fallibility of arguments in their premises, such as assumption-based argumentation [7] and variants of classical argumentation [6,13].

An important use of the *ASPIC*$^+$ framework is to investigate whether special cases or instances satisfy the four rationality postulates for argumentation-based inference proposed by [8]. To this end, [19,15] identify a number of jointly sufficient conditions under which these postulates are satisfied by special cases or instances of *ASPIC*$^+$, extending conditions identified by [8] for a special case of *ASPIC*$^+$ without preferences and without a knowledge base. Existing frameworks or systems can then be translated into *ASPIC*$^+$, and then investigated as to whether they satisfy these conditions. Likewise, new systems can be designed as instances of *ASPIC*$^+$ in such a way that they satisfy these conditions. Thus *ASPIC*$^+$ is not proposed as a framework superior to other approaches to argumentation. Rather, the aim of *ASPIC*$^+$ is to provide a framework in which to analyse existing and formalise new approaches, and identify conditions under which they satisfy rationality postulates (while possibly being extended with preferences over arguments).

Recently, it has been argued in [1] that the *ASPIC*$^+$ framework suffers from several weaknesses. We believe that the criticisms in [1] are based on a number of misconceptions on the nature of *ASPIC*$^+$. Among other things, our definitions and theorems are misread, our account of consistency is incompletely presented, claims that certain examples are counterexamples to our results are incorrect, claims are ascribed to us that we have not made, and more generally, the nature of *ASPIC*$^+$ as a framework rather than a particular system is not appreciated. We think it important that [1]'s criticisms are publicly rebutted, so that the research community can make an informed assessment of their quality.[1]

We first in Section 2 review the *ASPIC*$^+$ framework and comment in more detail on its nature and possible uses. We then in Section 3 present the criticisms of [1] and discuss why we think they are not justified.

2. Preliminaries

2.1. Abstract argumentation frameworks

An *argumentation framework* (*AF*) [10] is a tuple $(\mathcal{A}, \mathcal{C})$, where $\mathcal{C} \subseteq \mathcal{A} \times \mathcal{A}$ is a binary attack relation on the arguments $\mathcal{A}$. The status of arguments is then evaluated as follows:

Definition 1 Let $(\mathcal{A}, \mathcal{C})$ be a *AF*. Then $S \subseteq \mathcal{A}$ is then said to be *conflict free* iff $\forall X, Y \in S$, $(X, Y) \notin \mathcal{C}$. For any $X \in \mathcal{A}$, X is acceptable with respect to some $S \subseteq \mathcal{A}$ iff $\forall Y$ s.t. $(Y, X) \in \mathcal{C}$ implies $\exists Z \in S$ s.t. $(Z, Y) \in \mathcal{C}$. Let $S \subseteq \mathcal{A}$ be *conflict free*. Then:

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

2.2. The ASPIC$^+$ *framework for structured argumentation*

The *ASPIC*$^+$ framework [19,15] defines arguments, as in [22], as inference trees formed by applying strict or defeasible inference rules to premises that are well-formed formulae (wff) in some logical language. The distinction between two kinds of inference rules is taken from [17,18]. Informally, if an inference rule's antecedents are accepted, then if the rule is strict, its consequent must be accepted *no matter what*, while if the rule is defeasible, its consequent must be accepted *if there are no good reasons not to accept it*. Arguments can be attacked on their (non-axiom) premises and on their applications of defeasible inference rules. Some attacks succeed as *defeats*, which is partly determined by preferences. The acceptability status of arguments is then defined by applying any of [10]'s semantics for abstract argumentation frameworks to the resulting set of arguments with its defeat relation.

[1]Early May 2012 Amgoud put an "updated and extended" version online at `http://www.irit.fr/~Leila.Amgoud/`. However, we choose to comment only on the version as officially published in [1].

ASPIC$^+$ is not a system but a framework for specifying systems. It defines the notion of an abstract *argumentation system* (a notion adapted from [22]) as a structure consisting of a logical language $\mathcal{L}$ with a binary contrariness relation $^-$ and a naming convention n for defeasible rules, a set $\mathcal{R}$ consisting of two subsets $\mathcal{R}_s$ and $\mathcal{R}_d$ of strict and defeasible inference rules, and a partial preorder $\leq$ on $\mathcal{R}_d$. *ASPIC*$^+$ as a framework does not make any assumptions on how these elements are defined in a given argumentation system, except for some minimal assumptions on $\leq$ (the idea to abstract from the precise nature of $\mathcal{L}/\mathcal{R}$ is taken from [22] while the idea to abstract from $^-$ and n is taken from [7] and [18], respectively). In many instantiations of *ASPIC*$^+$ the set of strict rules will be determined by the choice of the logical language $\mathcal{L}$: its formal semantics will tell which inference rules over $\mathcal{L}$ are valid and can therefore be added to $\mathcal{R}_s$.

The formal definitions of *ASPIC*$^+$ differ on minor points between the various publications. Unless specified otherwise, we below present the version of [15].

Definition 2 An **ASPIC**$^+$***argumentation system*** is a tuple $AS = (\mathcal{L}, ^-, \mathcal{R}, n, \leq)$ where:

- $\mathcal{L}$ is a logical language.
- $^-$ is a contrariness 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 $\varphi \in \overline{\psi}$, $\psi \in \overline{\varphi}$.
- $\mathcal{R} = \mathcal{R}_s \cup \mathcal{R}_d$ 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 such that $\mathcal{R}_s \cap \mathcal{R}_d = \emptyset$.
- $n : \mathcal{R}_d \longrightarrow \mathcal{L}$ is a naming convention for defeasible rules.
- $\leq$ is a partial preorder on $\mathcal{R}_d$.

We say that $^-$ *corresponds to negation* iff $\mathcal{L}$ contains a connective $\neg$ such that $\psi \in \overline{\varphi}$ just in case $\psi = \neg\varphi$ or $\varphi = \neg\psi$.

In the previous publications on *ASPIC*$^+$ the idea of a naming convention n was instead informally introduced when defining undercutting attack (see Definition 6 below). Informally, $n(r)$ is a wff in $\mathcal{L}$ which says that rule $r \in \mathcal{R}$ is applicable.

Definition 3 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.

This definition is generalised from [8], in which these two notions of consistency were defined for the special case where $^-$ corresponds to negation. Moreover, in [8], and also in [19,15], only direct consistency was explicitly defined (and called "consistency"), while indirect consistency was implicitly defined in the definition of the rationality postulate of indirect consistency. Note that the definition of indirect consistency is in [8] parametrised by the choice of $\mathcal{R}_s$ and in *ASPIC*$^+$ also by the choice of $^-$.

Definition 4 An **ASPIC**$^+$***knowledge base*** in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, n, \leq)$ is a pair $(\mathcal{K}, \leq')$ where:

- $\mathcal{K} \subseteq \mathcal{L}$, and $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_p \cup \mathcal{K}_a$ where these subsets of $\mathcal{K}$ are disjoint, and: $\mathcal{K}_n$ is the (necessary) *axioms* ; $\mathcal{K}_p$ is the *ordinary premises*; $\mathcal{K}_a$ is the *assumptions*.
- $\leq'$ is a partial preorder on the non-axiom premises $\mathcal{K} \setminus K_n$.

Intuitively, axiomatic premises cannot be attacked, the success of attacks (as defeats) on ordinary premises is contingent on preferences, while attacks on assumptions always result in defeats (*cf.* assumptions in [7]).

Arguments are defined as in [22], where for any argument A, $\texttt{Prem}$ returns all the formulas of $\mathcal{K}$ (*premises*) used to build A, $\texttt{Conc}$ returns A's conclusion, $\texttt{Sub}$ returns all of A's sub-arguments, $\texttt{Rules}$ and $\texttt{DefRules}$ respectively return all rules and all defeasible rules in A, and $\texttt{TopRule}(A)$ returns the last rule applied in A.

Definition 5 An **ASPIC**$^+$***argument*** A on the basis of a knowledge base $(\mathcal{K}, \leq')$ in an argumentation system $(\mathcal{L}, ^-, \mathcal{R}, n, \leq)$ is:

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

Furthermore, 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 $\texttt{Prem}(A) \not\subseteq \mathcal{K}_n$.

Note that classical-logic approaches to argumentation (e.g., [4,13]) require that the premises of arguments are consistent. In order to formalise these approaches as instances of *ASPIC*$^+$, we in [15] defined an argument A to be *c-consistent* if $Cl(\texttt{Prem}(A))$ does not contain contradictory conclusions.

Three kinds of *attack* are defined for *ASPIC*$^+$ arguments. B can attack A by attacking a premise (undermining attack) or a conclusion (rebutting attack) of A, or an inference step in A (undercutting attack). Rebutting and undercutting attack are only possible on applications of defeasible inference rules. This idea and the distinction between rebutting and undercutting attack are taken from [17,18]. Some kinds of attack succeed as *defeats* independently of preferences over arguments, whereas others succeed only if the attacked argument is not stronger than the attacking argument. The partial preorders on defeasible rules and non-axiom premises may be used in defining an ordering $\preceq$ on the constructed arguments (we assume the strict counterpart $\prec$ of $\preceq$). For example, [19] presents definitions of $\preceq$ according to the weakest and last link principles.

Definition 6 A **attacks** B iff A *undercuts*, *rebuts* or *undermines* B, where:
- A *undercuts* argument B (on B') iff $\texttt{Conc}(A) \in \overline{n(r)}$ for some $B' \in \texttt{Sub}(B)$ such that B''s top rule r is defeasible.[2]

[2]Note that in [19,15] we instead wrote "$\texttt{Conc}(A) \in \overline{B'}$ for some $B' \in \texttt{Sub}(B)$" and we stated that this implicitly assumes a naming convention for applications of defeasible rules in $\mathcal{L}$.

• A *rebuts* argument B (on B') iff $\texttt{Conc}(A) \in \overline{\varphi}$ for some $B' \in \texttt{Sub}(B)$ of the form $B_1'', \ldots, B_n'' \Rightarrow \varphi$. In such a case A *contrary-rebuts* B iff $\texttt{Conc}(A)$ is a contrary of φ.
• Argument A *undermines* B (on B') iff $\texttt{Conc}(A) \in \overline{\varphi}$ for some $B' = \varphi$, $\varphi \in \texttt{Prem}(B) \setminus \mathcal{K}_n$. In such a case A *contrary-undermines* B iff $\texttt{Conc}(A)$ is a contrary of φ or if $\varphi \in \mathcal{K}_a$.

An undercut, contrary-rebut, or contrary-undermine attack is said to be *preference-independent*, otherwise an attack is *preference-dependent*.

A **defeats** B iff (1) A attacks B on B' and (2) if A's attack on B' is *preference-dependent* then $A \not\prec B'$.

The success of rebutting and undermining attacks thus involves comparing the conflicting arguments at the points where they conflict. The definition of successful undermining exploits the fact that an argument premise is also a subargument. The rationale for distinguishing preference-independent attacks is discussed in detail in [21], where it is argued that these attacks already embody a preference for the attacking over the attacked argument (e.g. attacks on negation-as-failure assumptions in logic programming which can be modelled as contrary-undermining attacks in *ASPIC*$^+$).

Adding an argument ordering to an argumentation system and a knowledge base yields an argumentation theory, which combined with the attack relation induces a structured argumentation framework. This framework in turn induces a Dung abstract argumentation framework.

Definition 7 An ***argumentation theory*** is a tuple $AT = (AS, KB)$ where AS is an argumentation system and KB is a knowledge base in AS.

A ***structured argumentation framework (*****SAF*****)*** defined by AT, is a triple $\langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ where $\mathcal{A}$ is the set of all (or all c-consistent) arguments constructed from KB in AS, $\preceq$ is a partial preorder on $\mathcal{A}$, and $(X, Y) \in \mathcal{C}$ iff X attacks Y.

Let $\Delta = \langle \mathcal{A}, \mathcal{C}, \preceq \rangle$ be a *SAF*, and $\mathcal{D} \subseteq \mathcal{A} \times \mathcal{A}$, where $(X, Y) \in \mathcal{D}$ iff X defeats Y. The ***extensions of the SAF*** Δ are the extensions of the Dung framework $(\mathcal{A}, \mathcal{D})$, as defined in Definition 1, with $\mathcal{C}$ replaced by $\mathcal{D}$.

Note that in [15], a variation of the *ASPIC*$^+$ framework is considered in which the conflict-freeness of a set of arguments is defined with respect to attack instead of defeat. For present purposes this variation is irrelevant and will henceforth be ignored.

Let us now formally state [8]'s rationality postulates for *ASPIC*$^+$. For any set S of arguments, let $\texttt{Conc}(S)$ be the set of all conclusions of any argument in S. Then for any $\Delta = (\mathcal{A}, \mathcal{C}, \preceq)$:

- Δ satisfies **subargument closure** iff for any complete extension E of Δ it holds that if $A \in E$, then $\texttt{Sub}(A) \subseteq E$.
- Δ satisfies **direct consistency** iff for any complete extension E of Δ it holds that $\texttt{Conc}(E)$ is directly consistent.
- Δ satisfies **indirect consistency** iff for any complete extension E of Δ it holds that $Cl(\texttt{Conc}(E))$ is directly consistent.
- Δ satisfies **closure under strict rules** iff for any complete extension E of Δ it holds that $\texttt{Conc}(E) = Cl(\texttt{Conc}(E))$.

In [8] the second and third postulate say "consistent" instead of "directly consistent" but consistency is in [8] still defined in the same way as in our Definition 3 of direct consistency. So these formulations of the postulates are equivalent to those of [8] except that in *ASPIC*$^+$ the notion of direct consistency is not defined in terms of negation but is generalised to arbitrary contrariness functions. This difference does not play a role in [1]'s criticisms.

2.3. *On the nature and uses of* ASPIC$^+$

It is important to realise that *ASPIC*$^+$ is not a system but a framework for specifying systems, such that these systems can be analysed on their properties, for instance, on whether they satisfy the above four rationality postulates. In future research we also want to address [9]'s recently proposed additional postulates of so-called 'non-interference' and 'crash resistance'.

One consequence of the framework-nature of *ASPIC*$^+$ is that it cannot be criticised by showing that it allows for instantiations that violate these postulates, since it is the very aim of *ASPIC*$^+$ to enable investigations of *whether* instantiations satisfy the postulates. Moreover, in [19,15] much guidance is given to users of the *ASPIC*$^+$ framework to ensure that their instantiations satisfy the postulates, in the form of a set of jointly sufficient conditions for their satisfaction, extending those identified in [8]. These conditions state some intuitively rational conditions on the strict rules and axiom premises (e.g., that strict rules admit contrapositive reasoning and the axiom premises are indirectly consistent), and that the argument ordering $\preceq$ is 'reasonable' in that it respects, for example, that strict and firm arguments are stronger than plausible and/or defeasible arguments. This guidance can also be used to investigate existing systems on their properties, namely, by translating them as an instance or special case of *ASPIC*$^+$ and by then investigating whether they satisfy our sufficient conditions. For example:

- In [19] assumption-based argumentation was shown to be a special case of *ASPIC*$^+$ with only strict inference rules, only assumption-type premises and no preferences. Because of this result, the sufficient conditions identified in [8] and [19] for satisfying [8]'s consistency postulates also apply to assumption-based argumentation, which in general does not satisfy these postulates.
- In [15] two forms of classical argumentation were shown to be a special case of *ASPIC*$^+$ with a propositional language, with only ordinary premises, with as strict rules all propositionally valid inferences and with no defeasible rules, and with all arguments being c-consistent. Then [19]'s weakest-link argument ordering was used to yield a preference-based version of classical argumentation that satisfies [8]'s rationality postulates.
- In [20] the Carneades system of [12] was shown to be an instance of *ASPIC*$^+$ with no defeat cycles, so that Carneades theories always have a unique extension in any of Dung's semantics (this result requires the use of so-called issue premises as included in *ASPIC*$^+$ in [19]). Moreover, in [21] the translation was exploited to show that Carneades satisfies all four rationality postulates of [8].
- Finally, in [16] it is shown that instantiations of *ASPIC*$^+$ in which the set of strict rules $\mathcal{R}_s$ is derived from a Tarskian abstract logic, and where preferences over arguments may be included, are well-behaved with respect to [8]'s rationality postulates (confirming informal conjectures made in [15]).

3. Amgoud's criticisms of *ASPIC*$^+$

3.1. ASPIC$^+$ *on consistency*

[1] claims that *ASPIC*$^+$'s notion of consistency (above called direct consistency) is too restricted and leads to a number of problems. However, here the *ASPIC*$^+$ framework is incompletely presented, since as explained above, it also contains the notion of indirect consistency (taken from [8]), and in [19,15] results are proven on [8]'s postulates for both direct and indirect consistency. (In fact, indirect consistency immediately follows from the proofs of direct consistency and closure under strict rules.)

3.2. Alleged counterexamples to consistency results for ASPIC$^+$

[1] goes on to present three examples claimed to be counterexamples to our theorems in [19,15] which state satisfaction of [8]'s consistency postulates. We now show that these claims are false. To start with, we note that all three purported examples satisfy the conditions under which our theorems hold, so if the extensions in the three examples are inconsistent, then [1] is correct in saying that the examples are counterexamples to our theorems. However, it can be shown that all extensions in the examples are both directly and indirectly consistent.

Example 1 Consider Example 1 in [1], in which it is assumed that $\mathcal{L}$ contains propositional formulas and that $X = \{x, x \supset y, \neg y\}$[3] and that $\overline{x} = \{\neg x, \neg\neg\neg x, \ldots\}$, $\overline{\neg y} = \{y, \neg\neg y, \ldots\}$ and $\overline{(x \supset y)} = \{x \wedge \neg y, \neg\neg(x \wedge \neg y), \ldots\}$. Amgoud sets $\mathcal{K}_p = X$ and leaves all other sets, including $\mathcal{R}$, empty.

Since $\mathcal{R}$ is empty, the only arguments that can be constructed are the elements of X (i.e., each premise in $\mathcal{K}_p = X$ is an argument). Moreover, according to the given contrariness relation, no two elements in X are a contrary or contradictory of the other, and so no argument is attacked, and so in all of [10]'s semantics there is a unique extension, namely X. Then, since $\mathcal{R}$ is empty, $\texttt{Conc}(X) = X$. So we must verify whether X is directly or indirectly inconsistent.

[1] correctly observes that according to *ASPIC*$^+$'s definition of (direct) consistency the set X is consistent but then she adds "whereas it is not". Apparently, when saying the latter, $\mathcal{L}$ is being interpreted as a standard propositional language and *ASPIC*$^+$'s notion of consistency is accordingly being replaced with standard propositional-logic's notion of consistency. However, it is crucial to note that [8]'s rationality postulates of direct and indirect consistency do not assume the standard propositional notion of consistency but instead the notions of direct and indirect consistency as defined in [8] and above in Definition 3. Now, as is indeed observed in [1], the set X is directly consistent, while to determine whether X is indirectly consistent, not the standard propositional notion of consistency must be used but instead [8]'s notion of indirect consistency as parametrised by the choice of strict rules $\mathcal{R}_s$. And since in [1]'s example $\mathcal{R}_s$ is empty, $Cl(X) = X$ and then it follows that X is also indirectly consistent. This example is therefore not a counterexample to any of our results on [8]'s rationality postulates.

[3] [1] in fact uses the symbol $\rightarrow$ here. In the context of this example $\rightarrow$ is apparently interpreted as material implication in the language $\mathcal{L}$. Therefore, to disambiguate from the use of $\rightarrow$ in strict inference rules, we write the material implication as $\supset$.

At first sight, it would seem that the only reason why this is not a counterexample to our consistency results is that *ASPIC*$^+$'s notions of consistency are non-standard (since X is clearly inconsistent if $\mathcal{L}$ is interpreted as in standard propositional logic and if consistency is defined as standard propositional consistency). However, here it is crucial to note that there is no reason whatsoever why $\mathcal{L}$ *should* be interpreted according to standard propositional logic. As is well known in logic, a given logical language can be interpreted in many different ways. For example, the language of Amgoud's example can be interpreted as in standard propositional logic but also as in intuitionistic logic, paraconsistent logics, relevant logics, many-valued logics, and so on. This also means that a given logical language does not come with a fixed notion of consistency. For this reason *ASPIC*$^+$ as a general framework cannot assume a fixed interpretation of a given logical language and must, as in [8], parametrise the notion of indirect consistency with the choice of strict rules. The fact that *ASPIC*$^+$ thus allows for non-standard notions of consistency is not a weakness of *ASPIC*$^+$ but a strength, since this makes a wide range of alternative logical instantiations of *ASPIC*$^+$ possible.

Moreover, it is very well possible in *ASPIC*$^+$ to interpret Amgoud's choice of $\mathcal{L}$ in such a way that *ASPIC*$^+$'s notion of indirect consistency coincides with standard propositional consistency. One way to do so is to let $^-$ correspond to negation and to let $\mathcal{R}_s$ consist of all inference rules over $\mathcal{L}$ that are valid in standard propositional logic. Let $S \vdash \varphi$ mean that there is a strict argument for φ constructible from the premises in S and let $S \vdash_{PL} \varphi$ denote that S logically implies φ in standard propositional logic. Then let 'S is PL-consistent' mean that for no φ it holds that $S \vdash_{PL} \varphi$ and $S \vdash_{PL} \neg\varphi$. Then

Theorem 1 *Let $\mathcal{L}$ be any propositional language such that $^-$ corresponds to negation and $S \rightarrow \varphi \in \mathcal{R}_s$ iff $S \vdash_{PL} \varphi$. Then S is indirectly consistent iff S is PL-consistent.*

PROOF. (Sketch) It suffices to prove for any $S \subseteq \mathcal{L}$ and any $\varphi \in \mathcal{L}$ that $S \vdash \varphi$ iff $S \vdash_{PL} \varphi$. From right to left is immediate from the choice of $\mathcal{R}_s$, while from left to right is proven with induction on the construction of a strict argument for φ from S. QED

Let us illustrate this theorem by modifying [1]'s example, letting $^-$ correspond to negation and letting $\mathcal{R}_s$ be $\{S \rightarrow \varphi \mid S \vdash_{PL} \varphi\}$. Then an infinite number of further arguments can be constructed, including the following arguments that undermine-attack any element of X, namely

$A = \neg y, x \supset y \rightarrow \neg x$
$B = \neg y, x \rightarrow \neg(x \supset y)$
$C = x, x \supset y \rightarrow y$

By conflict-freeness of extensions, at most two of these three arguments can be in the same extension. Moreover, it immediately follows from Theorem 32 of [16, p. 45] that there are three preferred-and-stable extensions, having the following conclusion sets

$\texttt{Conc}(\texttt{E}_1) = Cl(\{x, \neg y\})$
$\texttt{Conc}(\texttt{E}_2) = Cl(\{x, x \supset y\})$
$\texttt{Conc}(\texttt{E}_3) = Cl(\{\neg y, x \supset y\})$

which are all three both indirectly consistent and PL-consistent. Finally, it can be shown that there is one grounded extension, with the conclusion set

$\texttt{Conc}(\texttt{E}_4) = Cl(\emptyset)$

which is also both indirectly consistent and PL-consistent.

Exactly the same analysis applies to [1]'s second alleged counterexample:

Example 2 [1] again assumes that $\mathcal{L}$ contains propositional formulas, and sets $\mathcal{R}_d = \{\Rightarrow x; \Rightarrow \neg x \vee y; \Rightarrow \neg y\}$. All other sets are empty (including $\mathcal{R}_s$) and apparently it is assumed that the contrariness relation corresponds to negation.

This example has a single complete extension E, in which all three defeasible rules are applied, so $\texttt{Conc}(E) = \{x, \neg x \vee y, \neg y\}$. Again [1] claims that this is a counterexample to our consistency results but again this is false, for the same reasons as above. Moreover, once again, if $\mathcal{R}_s$ is chosen to consist of all valid propositional inferences, then the result coincides with standard propositional logic. We then obtain four complete extensions: in three of them (which are stable and preferred) two of the defeasible rules are applied and in the fourth (the grounded extension) none of them is applied. And again all extensions are both indirectly consistent and PL-consistent.

[1] finally claims to have a counterexample to our results in [15] stating that if *ASPIC*$^+$ is instantiated with classical logic, then (if the axiom premises are indirectly consistent and the argument ordering is reasonable) all four of [8]'s rationality postulates are satisfied. This claim is also false, since in the alleged counterexample as constructed in [1] several arguments rebut on the conclusions of *strict* inferences in other arguments, which is explicitly precluded by *ASPIC*$^+$'s Definition 6 of attack.

3.3. Domain-specific inference rules

As explained in Section 4 of [19], the inference rules from $\mathcal{R}_s$ and $\mathcal{R}_d$ can be used in two ways. They can be used to express domain-specific knowledge, such as 'birds fly' or 'all penguins are birds', or they can be used to express general patterns of reasoning such as the valid inferences of classical logic ($\mathcal{R}_s$), Pollock's [18] principles of epistemic cognition ($\mathcal{R}_d$) or argumentation schemes ($\mathcal{R}_d$). However, in both uses their logical role remains the same. Quoting from [19, p. 104]:

> The inference rules of argumentation systems are not part of the logical language $\mathcal{L}$ but are metalevel constructs.

This does not change if the inference rules are used in a domain-specific way. Yet [1, pg. 2] thinks that by using *ASPIC*$^+$'s inference rules to encode domain-specific information, they somehow move from $\mathcal{R}$ to $\mathcal{L}$:

> Prakken claims that strict and defeasible rules may play two roles: either they encode information of the knowledge base, in which case they are part of the language $\mathcal{L}$, or they represent inference rules, in which case they are part of the consequence operator ...

This is another misreading of our definitions. We therefore need not discuss Section 4 of [1], in which *ASPIC*$^+$'s inference rules are assumed to be part of $\mathcal{L}$.

3.4. ASPIC$^+$ *and abstract logics*

Amgoud next discusses *ASPIC*$^+$ in the light of [2,3]'s logical framework for argumentation. This framework is built on Tarski's notion of an abstract logic, which is a pair $(\mathcal{L}, Cn)$, where $\mathcal{L}$ is a language and the consequence operator Cn is a function from $2^{\mathcal{L}}$ to $2^{\mathcal{L}}$ satisfying the following conditions for all $X \subseteq \mathcal{L}$:

1. $X \subseteq Cn(X)$
2. $Cn(Cn(X)) = Cn(X)$
3. $Cn(X) = \bigcup_{Y \subseteq_f X} Cn(Y)$
4. $Cn(\{p\}) = \mathcal{L}$ for some $p \in \mathcal{L}$
5. $\mathrm{Cn}(\emptyset) \neq \mathcal{L}$

Here $Y \subseteq_f X$ means that Y is a finite subset of X. A set $X \subseteq \mathcal{L}$ is defined as *consistent* if $Cn(X) \neq \mathcal{L}$. [2,3] then define arguments and various kinds of attack relations, and investigate consistency properties of various types of attack relations under Dung's [10] semantics.

In [1] the claim is ascribed to [15] that *ASPIC*$^+$ "captures even Tarskian monotonic logics". What is apparently meant by this that we would claim that the 'logic' as captured by the sets $\mathcal{R}_s$ and $\mathcal{R}_d$ is Tarskian. [1] then attempts to prove that this is not true. However, the claim ascribed to [15] cannot be found in that paper. As noted above at the end of Section 2, we do in [15] informally make a number of other claims about the relation between *ASPIC*$^+$ and abstract logics and we have meanwhile formally proved these claims in [16].

In discussing these issues, [1] defines two "possible" *ASPIC*$^+$-like Cn operators and then proves that neither of them is Tarskian. However, it is not shown in [1] that these two Cn operators correspond to anything in *ASPIC*$^+$, so strictly speaking it is not yet known whether [1]'s results are relevant for *ASPIC*$^+$. Nevertheless, [1]'s claim is true for an obvious definition of Cn that is clearly related to *ASPIC*$^+$, namely:

- $p \in Cn(X)$ iff there exists an *ASPIC*$^+$ argument A, with $\texttt{Conc}(A) = p$ and $\texttt{Prem}(A) = X$.

With this definition of Cn, condition (4) of the definition of an abstract logic is in general not satisfied. Consider an AT with $\mathcal{K}_p = \{p\}$, $R_s = \emptyset$ and $R_d = \{p \Rightarrow q\}$. Also, condition (5) is not in general satisfied. Consider any AT with $\mathcal{K} = \emptyset$ and $R_d = \{\Rightarrow p \mid p \in \mathcal{L}\}$. So if the 'logic' as expressed in the sets of inference rules $\mathcal{R}_s$ and $\mathcal{R}_d$ is equated with the existence of an *ASPIC*$^+$ argument, then [1] is right in claiming that *ASPIC*$^+$'s underlying 'logic' is in general not Tarskian. However, recall that we never claimed the opposite.

Moreover, we do not regard this as a flaw of the *ASPIC*$^+$ framework. Instances of *ASPIC*$^+$ can be defined that are not Tarskian but that still satisfy all rationality postulates. Choose, for example, $\mathcal{L}$ to be a set of propositional or first-order literals as in e.g. Defeasible Logic Programming [11]. This instantiation may, depending on the choice of strict and defeasible rules, not correspond to a Tarskian abstract logic. Yet if all strict rules are transposed and any reasonable argument ordering is chosen, our results imply that all four of [8]'s postulates are satisfied.

Finally, we note that [2,3]'s abstract-logic approach like *ASPIC*$^+$ allows for instantiations in which its notion of consistency deviates from that of classical logic. Consider a language $\mathcal{L}$ closed under $\neg$ and with a distinguished element $\bot$, and define $Cn(S) = S$ for every $S \subseteq \mathcal{L}$ that does not contain $\bot$ and $Cn(S) = \mathcal{L}$ otherwise. Such a Cn is Tarskian but is such that for every $\varphi \in \mathcal{L}$ that does not equal $\bot$ we have that $\{\varphi, \neg\varphi\}$ is consistent in the sense of the abstract-logic approach. We do not regard this as a flaw of [2,3]'s abstract-logic approach, since just as *ASPIC*$^+$ it is not a particular system but a general framework. However, if in [1] *ASPIC*$^+$ is criticised for allowing non-standard interpretations of consistency, then the question arises why [2,3]'s abstract-logic approach is not criticised for the same reason.

3.5. *Some other misconceptions*

We briefly discuss some other misconceptions in [1]. First, it is claimed that in *ASPIC*$^+$ preferences can only be applied to symmetric attacks, but this is not true. Consider an argument A with a strict top rule for p and an argument B with a defeasible top rule for $\neg p$ and let the contrariness relation correspond to negation. Then A asymmetrically rebuts B but preferences can be applied to this attack.

Next, [1] claims that our condition on the argument ordering being reasonable (which among other things requires that strict-and-firm arguments are strictly preferred to plausible or defeasible arguments) is limiting. In particular, [1] notes that *ASPIC*$^+$ cannot capture instances of [5]'s value-based argumentation frameworks (VAFs) in which an argument that is plausible or defeasible promotes a more important value than an argument that is strict and firm. We claim that, if such an instantiation of VAFs can be realistically given at all (which we doubt), then this is a flaw of the VAF instantiation and not of *ASPIC*$^+$.

Then [1] gives another example (her Example 3) with an alleged counterintuitive outcome, namely, an argumentation system allowing the following two arguments:

A: $\Rightarrow p, p \Rightarrow q, q \Rightarrow r, r \rightarrow x$
B: $\rightarrow d, d \rightarrow e, e \rightarrow f, f \Rightarrow \neg x$

[1] correctly observes that A asymmetrically attacks B so that x rather than $\neg x$ is concluded, which [1] regards as counterintuitive since A applies many more defeasible rules than B. Here the framework-nature of *ASPIC*$^+$ is relevant, which as said above, implies that *ASPIC*$^+$ as a framework cannot be criticised by giving problematic instances. In particular, in this example the strict rules are *not* closed under transposition or contraposition, which are identified in [19,15] as conditions for satisfying the consistency postulates. If the transpositions of all strict rules are added to $\mathcal{R}_s$, then the outcome criticised in [1] is not obtained, since then B can be extended with $\neg x \rightarrow \neg r$ (the transposition of $r \rightarrow x$) to an argument that rebuts A on its subargument for r.

[1] further claims that our definition of contrary-rebuts in Definition 6 is flawed since it would mean that if an argument A contrary-rebuts B on B's final conclusion, then B cannot contrary-rebut A on A's final conclusion (and so A asymmetrically attacks B). But the reason why [1] thinks this is flawed, is because she appeals to an *alternative* definition of contraries and contradictories that is taken from *some other* paper, and which does not allow for such asymmetry.

[1] then criticises *ASPIC*$^+$ for blurring the relation between object and metalanguage in the definition of undercutting attack. However, in [15] we say: "To model undercutting attacks on inferences, it is assumed that applications of inference rules can be expressed in the object language; the precise nature of this naming convention will be left implicit." (a similar phrase is in [19]). Incidentally, as mentioned above, we have meanwhile incorporated the naming convention in Definition 2 to avoid possible confusion.

Finally, [1] claims that *ASPIC*$^+$ would be at odds with [14]'s principle of right weakening of defeasible consequence, which would require that extensions are closed under defeasible rules. However, here she misreads [14]: their principle of right weakening instead says that defeasible consequences should be closed under classical entailment, which corresponds to [8]'s rationality postulate of closure under *strict* rules. Note also that in general extensions should not be closed under *defeasible* rules.

4. Conclusions

In this paper we have shown that Amgoud's criticisms of the *ASPIC*$^+$ framework are based on a number of misconceptions on the nature of *ASPIC*$^+$. We have shown that in [1] our account of consistency is incompletely presented, that our definitions and theorems are misread, that claims that certain examples are counterexamples to our results are incorrect, that claims are ascribed to us that we have not made, and more generally, that the nature of *ASPIC*$^+$ as a framework rather than a particular system is not appreciated. We sincerely hope that the research community will take our rebuttal into account when assessing whether [1]'s criticism of the *ASPIC*$^+$ framework is justified.

References

[1] L. Amgoud. Five weaknesses of ASPIC+. In *Proceedings of the 14th International Conference on Information Processing and Managing of Uncertainty in Knowledge-Based Systems*. LNAI, Springer Verlag, 2012.

[2] L. Amgoud and P. Besnard. Bridging the gap between abstract argumentation systems and logic. In *Proc. 3rd International Conference on Scalable Uncertainty (SUM'09)*, pages 12–27, 2009.

[3] L. Amgoud and P. Besnard. A formal analysis of logic-based argumentation systems. In *Proc. 4th International Conference on Scalable Uncertainty (SUM'10)*, pages 42–55, 2010.

[4] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1-3):197–215, 2002.

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

[6] P. Besnard and A. Hunter. *Elements of Argumentation*. MIT Press, 2008.

[7] 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.

[8] M. Caminada and L. Amgoud. On the evaluation of argumentation formalisms. *Artificial Intelligence*, 171(5-6):286–310, 2007.

[9] M. Caminada, W. Carnielli, and P.E. Dunne. Semi-stable semantics. *Journal of Logic and Computation*, to appear.

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

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

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

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

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

[15] S. Modgil and H. Prakken. Revisiting preferences and argumentation. In *IJCAI 2011*, pp. 1021–1026.

[16] S. Modgil and H. Prakken. A general account of argumentation with preferences: Technical report. *http://www.dcs.kcl.ac.uk/staff/smodgil/GAP.pdf*, 2012.

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

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

[19] H. Prakken. An abstract framework for argumentation with structured arguments. *Argument and Computation*, 1(2):93–124, 2010.

[20] 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.

[21] B. van Gijzel and H. Prakken. Relating Carneades with abstract argumentation via the ASPIC+ framework for structured argumentation. *Argument and Computation*, 1:21–47, 2012.

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