Persistence and Monotony Properties of Argumentation Semantics

Tjitze Rienstra[1] Chiaki Sakama[2] Leendert van der Torre[3]

[1] Interdisciplinary Centre for Security, Reliability and Trust
University of Luxembourg (tjitze@gmail.com)

[2] Wakayama University (sakama@sys.wakayama-u.ac.jp)

[3] Computer Science and Communication
University of Luxembourg (leon.vandertorre@uni.lu)

Abstract. We study a number of properties concerning the behaviour of semantics for Dung style abstract argumentation when the argumentation framework changes. The properties are concerned with how the evaluation of an argumentation framework changes if an attack between two arguments is added or removed. The results provide insight into the behaviour of these semantics in a dynamic context.

1 Introduction

Argumentation is a process that usually involves a series of steps taken to achieve a particular end. However, an argumentation framework (AF, for short) represents only a static snapshot of this process. To consider dynamics of argumentation, we need to consider AFs that change, for example due to the addition of new arguments and attacks. In this paper we focus on the behaviour of semantics for argumentation in a dynamic context, by studying properties related to how the evaluation of an AF changes if the AF changes.

Our work extends the growing literature on the general problem of change in Dung style argumentation theory. This includes studies of strong equivalence [12], enforcing [3] and revision in argumentation [6, 10]. In particular, this paper extends the approach of Boella et al. [4, 5], who studied *refinement* and *abstraction* principles, which are conditions under which the evaluation of an AF remains unchanged when an attack is added or removed. While they only focussed on semantics that yield a single extension or labelling, we extend their approach by considering three types of properties that also apply to semantics that yield multiple extensions or labellings:

- XY **addition persistence**: a σ labelling of an AF F in which x is labelled X and y is labelled Y is still a σ labelling of F after adding an attack from x to y.
- XY **removal persistence**: a σ labelling of an AF F in which x is labelled X and y is labelled Y is still a σ labelling of F if removing the attack from x to y.
- XY **skeptical monotony**: if in all σ labellings of an AF F, x is labelled X and y is labelled Y, then adding an attack from x to y does not lead to new σ labellings.

We systematically check, for each combination of labels X and Y (relying on Caminada's three-valued labellings [7]) whether XY addition persistence, removal persistence and skeptical monotony are satisfied under the grounded, complete, preferred, stable and semi-stable semantics. Our results provide insight into the behaviour of these semantics in a dynamic context. For example, **OO** addition persistence reflects the principle that a point of view represented by a labelling L need not be revised due to the addition of an attack from x to y, if both x and y are labelled *out* in L. This is intuitive because the added attack does not introduce a conflict with respect to L in this case. We show that some combinations of X and Y yield counterintuitive properties that are, indeed, not satisfied under any of the semantics we consider. Other properties are intuitive, such as the above mentioned **OO** addition persistence, and are satisfied under some semantics but not under all.

The layout of this paper is as follows. In section 2 we recall the necessary definitions concerning AFs, extension-based semantics and labelling-based semantics. In sections 3, 4 and 5 we introduce the addition persistence, removal persistence and skeptical monotony properties and we check whether they are satisfied or not. In section 6 we discuss related work. In section 7 we conclude and discuss a number of directions for future research.

2 Preliminaries

Formally, an AF is a directed graph represented by a set A of *arguments* and a binary relation $\rightsquigarrow$ over A called the *attack relation* [11]. To simplify our discussion we assume that A is a finite subset of a set $\mathcal{U}$ called the *universe of arguments*.

Definition 1. *Let $\mathcal{U}$ be a set whose elements are called* arguments*. An* argumentation framework *(AF, for short) is a pair $F = (A, \rightsquigarrow)$ where A is a finite subset of $\mathcal{U}$ and $\rightsquigarrow \subseteq A \times A$. We denote by $\mathcal{F}$ the set of all AFs.*

Given an AF $(A, \rightsquigarrow)$ we say that an argument $a \in A$ *attacks* an argument $b \in A$ if and only if $(a, b) \in \rightsquigarrow$. Given an AF F we denote by x^- (resp. B^-) the set of arguments attacking x (resp. some $x \in B$) and by x^+ (resp. B^+) the set of arguments attacked by x (resp. some $x \in B$).

2.1 Extension-based Semantics

An extension-based semantics is defined by a condition that an extension (i.e., a set of arguments) must satisfy so that it represents a rationally acceptable set of arguments. Three basic conditions are *conflict-freeness*, *admissibility* and *completeness*. An extension E is conflict-free if it is not self-attacking, admissible if it defends all its members, and complete if it furthermore includes all arguments that it defends. The notion of defence used here is defined as follows.

Definition 2. *[11] Let $F = (A, \rightsquigarrow)$ be an AF. An extension $E \subseteq A$* defends *an argument $y \in A$ if and only if for all $x \in A$ such that $x \rightsquigarrow y$, there is a $z \in E$ such that $z \rightsquigarrow x$. We define $\mathcal{D}_F(E)$ by $\mathcal{D}_F(E) = \{x \in A \mid E \text{ defends } x\}$.*

Definition 3. *[11] Let $F = (A, \rightsquigarrow)$ be an AF. An extension $E \subseteq A$ is* conflict-free *iff for no $x, y \in A$ it holds that $x \rightsquigarrow y$;* admissible *iff E is conflict-free and $E \subseteq \mathcal{D}_F(E)$; and* complete *iff E is conflict-free and $E = \mathcal{D}_F(E)$.*

An extension-based semantics σ can be represented by a function $\mathcal{E}_\sigma : \mathcal{F} \to 2^{2^{\mathcal{U}}}$ satisfying $\mathcal{E}_\sigma((A, \rightsquigarrow)) \subseteq 2^A$. The following definition presents some of the most-used semantics, namely the complete (*co*), grounded (*gr*), preferred (*pr*), semi-stable (*ss*) and stable (*st*) semantics.

Definition 4. *Let $F = (A, \rightsquigarrow)$ be an AF.*

- $\mathcal{E}_{co}(F) = \{E \subseteq A \mid E \text{ is a complete extension of } F\}$
- $\mathcal{E}_{gr}(F) = \{E \in \mathcal{E}_{co}(F) \mid \nexists E' \in \mathcal{E}_{co}(F) \text{ s.t. } E' \subset E\}$
- $\mathcal{E}_{pr}(F) = \{E \in \mathcal{E}_{co}(F) \mid \nexists E' \in \mathcal{E}_{co}(F) \text{ s.t. } E \subset E'\}$
- $\mathcal{E}_{ss}(F) = \{E \in \mathcal{E}_{co}(F) \mid \nexists E' \in \mathcal{E}_{co}(F) \text{ s.t. } E \cup E^+ \subset E' \cup E'^+\}$
- $\mathcal{E}_{st}(F) = \{E \in \mathcal{E}_{co}(F) \mid E \cup E^+ = A\}$

The grounded extension can also be characterized by a fix point theory.

Proposition 1. *[11, Theorem 25] Given an AF F, the grounded extension of F coincides with the least fixed point of $\mathcal{D}_F$.*

2.2 Labelling-based Semantics

While an extension only captures the arguments that are accepted in a given position, a labelling assigns to each argument an *acceptance status*. This approach can be traced back to Pollock [13]. We follow Caminada [7], who defined the semantics considered here using three-valued labellings: **I** (accepted), **O** (rejected) and **U** (undecided). The benefit of labellings over extensions is that they permit us to distinguish not only arguments that are accepted and not accepted, but also those that are explicitly rejected and those that are undecided.

Definition 5. *A* labelling *of an AF $(A, \rightsquigarrow)$ is a function $L : A \to \{\mathbf{I}, \mathbf{O}, \mathbf{U}\}$. Given a label $l \in \{\mathbf{I}, \mathbf{O}, \mathbf{U}\}$ we define $L^{-1}(l)$ as $\{x \in A \mid L(x) = l\}$. Given an AF F, we let $\mathcal{L}(F)$ denote the set of all labellings of F.*

We also denote a labelling L by a set of pairs $\{(x_1, L(x_1)), \ldots, (x_n, L(x_n))\}$. Complete labellings are defined as follows.

Definition 6. *Let $F = (A, \rightsquigarrow)$ be an AF. A labelling $L \in \mathcal{L}(F)$ is* complete *if and only if for all $x \in A$, $L(x) = \mathbf{I}$ iff for all $y \in x^-$, $L(y) = \mathbf{O}$; and $L(x) = \mathbf{O}$ iff for some $y \in x^-$, $L(y) = \mathbf{I}$.*

The following definition introduces the labelling-based versions of the semantics presented in the previous section [8].

Definition 7. *Let $F = (A, \rightsquigarrow)$ be an AF.*

- $\mathcal{L}_{co}(F) = \{L \in \mathcal{L}(F) \mid L \text{ is a complete labelling of } F\}$
- $\mathcal{L}_{gr}(F) = \{L \in \mathcal{L}_{co}(F) \mid \nexists L' \in \mathcal{L}_{co}(F) \ s.t. L'^{-1}(\mathbf{I}) \subset L^{-1}(\mathbf{I})\}$
- $\mathcal{L}_{pr}(F) = \{L \in \mathcal{L}_{co}(F) \mid \nexists L' \in \mathcal{L}_{co}(F) \ s.t. L^{-1}(\mathbf{I}) \subset L'^{-1}(\mathbf{I})\}$
- $\mathcal{L}_{ss}(F) = \{L \in \mathcal{L}_{co}(F) \mid \nexists L' \in \mathcal{L}_{co}(F) \ s.t. L'^{-1}(\mathbf{U}) \subset L^{-1}(\mathbf{U})\}$
- $\mathcal{L}_{st}(F) = \{L \in \mathcal{L}_{co}(F) \mid L^{-1}(\mathbf{U}) = \emptyset\}$

The following proposition establishes a correspondence between extensions and labellings. It has been shown that this is a one-to-one mapping [8].

Proposition 2. *Let $F = (A, \rightsquigarrow)$ be an AF and let $\sigma \in \{co, gr, pr, ss, st\}$.*

- *If $L \in \mathcal{L}_\sigma(F)$ then $L^{-1}(\mathbf{I}) \in \mathcal{E}_\sigma(F)$.*
- *If $E \in \mathcal{E}_\sigma(F)$ then $L \in \mathcal{L}_\sigma(F)$, where L is defined by $L(x) = \mathbf{I}$, if $x \in E$, $L(x) = \mathbf{O}$, if $x \in E^+$ and $L(x) = \mathbf{U}$, otherwise.*

3 Addition Persistence Properties

We first consider *addition persistence*, defined with respect to two labels X and Y. We say that a semantics σ satisfies XY addition persistence whenever a σ labelling of an AF F in which x is labelled X and y is labelled Y is still a σ labelling of F after adding an attack from x to y. Formally:

Definition 8. *Let σ be a semantics and let $X, Y \in \{\mathbf{O}, \mathbf{I}, \mathbf{U}\}$. We say that σ satisfies XY addition persistence if and only if for all $(A, \rightsquigarrow) \in \mathcal{F}$ and $x, y \in A$, if $L \in \mathcal{L}_\sigma((A, \rightsquigarrow))$, $L(x) = X$ and $L(y) = Y$, then $L \in \mathcal{L}_\sigma((A, \rightsquigarrow \cup \{(x, y)\}))$.*

For some combinations of X and Y, the semantics we consider obviously do not satisfy XY addition persistence. These are **II**, **IU** and **UI** addition persistence. The reason is that no two arguments x and y such that x attacks y ever receive these combinations of labels in a complete labelling.

Proposition 3. *The grounded, complete, preferred and semi-stable semantics do not satisfy* **II**, **IU** *and* **UI** *addition persistence.*

Under the stable semantics, where arguments are never labelled **U**, all properties involving **U**-labelled arguments are trivially satisfied.

Proposition 4. *The stable semantics satisfies* **UO**, **UU**, **UI**, **UO** *and* **UI** *addition persistence but does not satisfy* **II** *addition persistence.*

For the remaining combinations of X and Y, XY addition persistence may be considered desirable. For example, **OO**, **OU** and **OI** addition persistence together reflect the principle that a point of view L need not be revised due to the addition of an attack from x to y, if x is labelled **O**. This is intuitive because the added attack does not in this case introduce a conflict with respect to L, as the argument from which the added attack originates is rejected. Similar considerations apply to **UO**, **UU** and **IO** addition persistence, which also concern combinations where the added attack does not introduce conflict.

We now turn to the question whether the semantics we consider satisfy these remaining properties. Let us start with the grounded and complete semantics. While the grounded semantics satisfies **OO**, **OU**, **UO**, **UU** and **IO** addition persistence, it does not satisfy **OI** addition persistence.

Theorem 1. *The grounded semantics satisfies* **OO***,* **OU***,* **UO***,* **UU** *and* **IO** *addition persistence but not* **OI** *addition persistence.*

Proof. Due to lack of space we only provide a proof sketch. Satisfaction can be proven by induction on the construction of the grounded extension of F as the fix point of $\mathcal{D}_F$. A counterexample for **OI** addition persistence is provided below.

The failure of **OI** addition persistence is due to the fact that the addition of an attack from an **O** to an **I** labelled argument may lead to a new complete labelling in which both arguments are **U**. This new labelling will become the new grounded labelling. This is demonstrated by the following example.

Example 1 (Failure of **OI** *addition persistence under the grounded semantics).* Consider an AF with two arguments a, b, where a attacks b. In the grounded labelling, a is labelled **I** and b is labelled **O**. If we add an attack from b to a the grounded labelling assigns **U** to a and b.

The complete semantics satisfies all properties satisfied by the grounded semantics but it also satisfies **OI** addition persistence.

Theorem 2. *The complete semantics satisfies* **OO***,* **OU***,* **OI***,* **UO***,* **UU** *and* **IO** *addition persistence.*

Proof. Follows easily from definition 6. We omit this proof due to lack of space.

The preferred semantics satisfies all properties satisfied by the complete semantics, except for **UU** addition persistence.

Theorem 3. *The preferred semantics satisfies* **OO***,* **OU***,* **OI***,* **IO** *and* **UO** *addition persistence but not* **UU** *addition persistence.*

Proof. Let $(A, \rightsquigarrow) \in \mathcal{F}$ and let $L \in \mathcal{L}_{pr}((A, \rightsquigarrow))$ be a labelling s.t. either $L(x) = \mathbf{O}$ or $L(y) = \mathbf{O}$. We prove that $L \in \mathcal{L}_{pr}((A, \rightsquigarrow \cup \{(x, y)\}))$. If $x \rightsquigarrow y$ we are done. For the case $x \not\rightsquigarrow y$, assume the contrary. We have $L \in \mathcal{L}_{co}((A, \rightsquigarrow \cup \{(x, y)\}))$ and thus there must be an $L' \in \mathcal{L}_{pr}((A, \rightsquigarrow \cup \{(x, y)\}))$ such that $L^{-1}(\mathbf{I}) \subset L'^{-1}(\mathbf{I})$. It then follows that either $L'(x) = \mathbf{O}$ or $L'(y) = \mathbf{O}$ and hence $L' \in \mathcal{L}_{co}((A, \rightsquigarrow))$. This contradicts $L \in \mathcal{L}_{pr}((A, \rightsquigarrow))$. Hence $L \in \mathcal{L}_{pr}((A, \rightsquigarrow \cup \{(x, y)\}))$. A counterexample for **UU** addition persistence is provided below.

Failure of **UU** addition persistence is due to the fact that an attack from a **U** to another **U** labelled argument may lead to a new complete labelling in which one of the two is labelled **I**. This new labelling will replace the old labelling as one of the preferred labellings. This is demonstrated by the following example.

Example 2 (Failure of **UU** *addition persistence under the preferred semantics).* Consider an AF with two arguments a, b, where a attacks b and a is self-attacking. The unique preferred labelling assigns **U** to a and b. If we add an attack from b to a the unique preferred labelling assigns **O** to a and **I** to b.

We already saw that addition persistence properties involving **U**-labelled arguments are trivially satisfied under the stable semantics. The remaining interesting properties are **OO**, **OI** and **IO** addition persistence. They are all satisfied.

Theorem 4. *Stable semantics satisfies* **OO***,* **OI** *and* **IO** *addition persistence.*

Proof. This follows from the proof of theorem 2 together with the fact that a labelling is stable if and only if it is complete and no argument is labelled **U**.

Under the semi-stable semantics, none of the remaining XY addition persistence properties are satisfied.

Theorem 5. *The semi-stable semantics does not satisfy* **IO***,* **UO***,* **OO***,* **OU** *or* **OI** *addition persistence.*

We list here the counterexamples.

Example 3 (Failure of **IO***,* **UO** *and* **OO** *addition persistence under semi-stable semantics).* The AF shown in figure 1 has a unique semi-stable labelling $L = \{(a, \mathbf{I}), (b, \mathbf{O}), (c, \mathbf{I}), (d, \mathbf{O}), (e, \mathbf{U}), (f, \mathbf{U})\}$. If we add an attack from a to b (labelled **I** and **O**); from f to b (labelled **U** and **O**); or from d to b (both labelled **O**), L is no longer a semi-stable labelling. Instead we get the unique semi-stable labelling $\{(a, \mathbf{I}), (b, \mathbf{O}), (c, \mathbf{O}), (d, \mathbf{I}), (e, \mathbf{O}), (f, \mathbf{I})\}$.

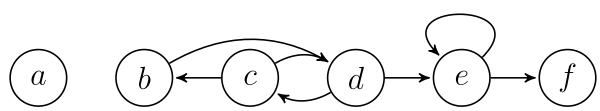

Fig. 1: Failure of **IO**, **UO** and **OO** addition persistence under semi-stable semantics.

Example 4 (Failure of **OU** *addition persistence under semi-stable semantics).* The AF shown in figure 2 has a unique semi-stable labelling $L = \{(a, \mathbf{U}), (b, \mathbf{I}), (c, \mathbf{O}), (d, \mathbf{U})\}$. If we add an attack from c to a (labelled **O** and **U**, respectively) then L is no longer a semi-stable labelling. Instead we get the unique semi-stable labelling $\{(a, \mathbf{O}), (b, \mathbf{O}), (c, \mathbf{I}), (d, \mathbf{O})\}$.

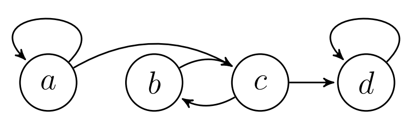

Fig. 2: Failure of **OU** addition persistence under semi-stable semantics.

Example 5 (Failure of **OI** *addition persistence under semi-stable semantics).* Consider an AF with three arguments a, b and c, where a attacks b, b attacks c and c is self-attacking. The unique semi-stable labelling is $\{(a, \mathbf{I}), (b, \mathbf{O}), (c, \mathbf{U})\}$. If we add an attack from b to a (labelled **O** and **I**, respectively) the unique semi-stable labelling becomes $\{(a, \mathbf{O}), (b, \mathbf{I}), (c, \mathbf{O})\}$.

Let us summarize the results obtained in this section. At the start we established that none of the semantics we consider satisfy **II**, **IU** and **UI** addition persistence, except the stable semantics, which does not satisfy **II** addition persistence but trivially satisfies all properties involving **U**-labelled arguments. The complete and stable semantics can be considered the best behaved ones, as they satisfy all remaining properties, while the semi-stable semantics can be considered the worst behaved one, as it satisfies none. Table 1 contains an overview of the results.

Grounded:

		X		
		O	**U**	**I**
Y	**O**	✓	✓	✓
	U	✓	✓	-
	I	-	-	-

Complete:

		X		
		O	**U**	**I**
Y	**O**	✓	✓	✓
	U	✓	✓	-
	I	✓	-	-

Preferred:

		X		
		O	**U**	**I**
Y	**O**	✓	✓	✓
	U	✓	-	-
	I	✓	-	-

Stable:

		X		
		O	**U**	**I**
Y	**O**	✓	✓	✓
	U	✓	✓	✓
	I	✓	✓	-

Semi-Stable:

		X		
		O	**U**	**I**
Y	**O**	-	-	-
	U	-	-	-
	I	-	-	-

Table 1: Overview of addition persistence properties.

4 Removal Persistence

We now consider the property of *removal persistence*. We say that a semantics σ satisfies XY removal persistence whenever every σ labelling of an AF F in which two arguments x and y are labelled X and Y, respectively, is still a σ labelling of F after removing the attack from x to y. Formally:

Definition 9. *Let σ be a semantics and let $X, Y \in \{\mathbf{O}, \mathbf{I}, \mathbf{U}\}$. We say that σ satisfies XY removal persistence if and only if for all $(A, \rightsquigarrow) \in \mathcal{F}$ and $x, y \in A$, if $L \in \mathcal{L}_\sigma((A, \rightsquigarrow))$, $L(x) = X$ and $L(y) = Y$, then $L \in \mathcal{L}_\sigma((A, \rightsquigarrow \setminus \{(x, y)\}))$.*

Like in the previous section we now determine, for all semantics that we consider, and for all combinations of X and Y, whether XY removal persistence is satisfied or not. We first establish a number of obvious cases. First of all, **II**, **IU** and **UI** removal persistence are trivially satisfied under all semantics we consider, because these combinations of labels are never assigned by a complete labelling to any two arguments x and y where x attacks y. Similarly, all removal properties involving **U**-labelled arguments are trivially satisfied under the stable labellings, where argument are never labelled **U**. Furthermore **IO** removal persistence fails under all semantics we consider, because in a complete labelling, an argument is labelled **O** only if some attacker is labelled **I**. Thus, removing an attack from an argument labelled **I** in L to an argument labelled **O** in L may invalidate L. The same holds for **UU** removal persistence, which fails under all semantics we consider, except under the stable semantics, where it is trivially satisfied.

Proposition 5. *The grounded, complete, preferred and semi-stable semantics satisfy* **II**, **IU** *and* **UI** *removal persistence but do not satisfy* **IO** *and* **UU** *removal persistence.*

Proposition 6. *The stable semantics satisfies* **II**, **UO**, **UU**, **UI**, **OU** *and* **IU** *removal persistence but does not satisfy* **IO** *removal persistence.*

In the remainder of this section we focus on **OO**, **OU**, **OI** and **UO** removal persistence. All these properties may be considered desirable. For example, **OO** removal persistence reflects the principle that a point of view L need not be revised due to the removal of an attack from x to y, when both x and y are labelled **O** in L. This is an intuitive principle, because y is not in this case rejected due to being attacked by x, and removing it does not change whether or not the rejection of y is justified. Similar considerations apply to **OU**, **OI** and **UO** removal persistence.

Let us start with the grounded, complete and preferred semantics.

Theorem 6. *The grounded, complete and preferred semantics satisfy* **OO**, **OU**, **OI** *and* **UO** *removal persistence.*

Proof. Grounded: Due to space constraints we only provide a sketch of the proof. The satisfied properties can be proven by induction on the construction of the grounded extension of an AF F as the fix point of $\mathcal{D}_F$. *Complete:* Follows easily from the definition of a complete labelling. *Preferred:* Let $(A, \rightsquigarrow) \in \mathcal{F}$ and let $L \in \mathcal{L}_{pr}((A, \rightsquigarrow))$ be a labelling s.t. $L(x) = \mathbf{O}$. We prove that $L \in \mathcal{L}_{pr}((A, \rightsquigarrow \setminus \{(x, y)\}))$. If $x \not\rightsquigarrow y$ we are done. For the case $x \rightsquigarrow y$, assume the contrary. We have $L \in \mathcal{L}_{co}((A, \rightsquigarrow \setminus \{(x, y)\}))$ and thus there must be an $L' \in \mathcal{L}_{pr}((A, \rightsquigarrow \setminus \{(x, y)\}))$ such that $L^{-1}(\mathbf{I}) \subset L'^{-1}(\mathbf{I})$. It then follows that $L'(x) = \mathbf{O}$ and hence $L' \in \mathcal{L}_{co}((A, \rightsquigarrow))$. This contradicts $L \in \mathcal{L}_{pr}((A, \rightsquigarrow))$. Hence $L \in \mathcal{L}_{pr}((A, \rightsquigarrow \setminus \{(x, y)\}))$. This proves that the preferred semantics satisfies **OO**, **OU** and **OI** persistence. The proof for **UO** is similar.

Let us move on to the stable semantics. Proposition 6 already lists a number of properties that are trivially satisfied. The following theorem concerns the two remaining non-trivial properties.

Theorem 7. *The stable semantics satisfies* **OO** *and* **OI** *removal persistence.*

Proof. This follows from theorem 6 together with the fact that a labelling is stable if and only if it is complete and no argument is labelled **U**.

Under the semi-stable semantics we see that all the remaining properties (**OO**, **OU**, **OI** and **UO** removal persistence) fail.

Example 6 (Failure of **OO** *and* **OI** *removal persistence under semi-stable semantics).* The AF shown in figure 3 has a unique semi-stable labelling $L = \{(a, \mathbf{O}), (b, \mathbf{I}), (c, \mathbf{O}), (d, \mathbf{I}), (e, \mathbf{O}), (f, \mathbf{U})\}$. If we remove the attack from a to e (both labelled **O**) then L is no longer a semi-stable labelling. Instead we get the unique semi-stable labelling $\{(a, \mathbf{I}), (b, \mathbf{O}), (c, \mathbf{I}), (d, \mathbf{O}), (e, \mathbf{I}), (f, \mathbf{O})\}$. Similarly, if we remove the attack from c to b (labelled **O** and **I**, respectively) then L is no longer semi-stable. Instead we get the unique semi-stable labelling $\{(a, \mathbf{O}), (b, \mathbf{I}), (c, \mathbf{I}), (d, \mathbf{O}), (e, \mathbf{I}), (f, \mathbf{O})\}$.

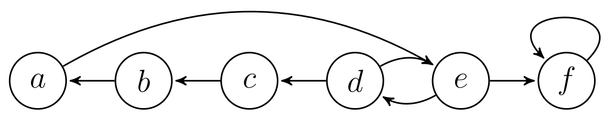

Fig. 3: Failure of **OO** and **OI** removal persistence under semi-stable semantics.

Example 7 (Failure of **OU** *removal persistence under semi-stable semantics).* The AF shown in figure 4 has a unique semi-stable labelling $L = \{(a, \mathbf{U}), (b, \mathbf{U}), (c, \mathbf{O}), (d, \mathbf{I}), (e, \mathbf{O}), (f, \mathbf{U})\}$. If we remove the attack from c to b (labelled **O** and **U**, respectively) then L is no longer a semi-stable labelling. Instead we get the unique semi-stable labelling $L = \{(a, \mathbf{O}), (b, \mathbf{I}), (c, \mathbf{I}), (d, \mathbf{O}), (e, \mathbf{I}), (f, \mathbf{O})\}$.

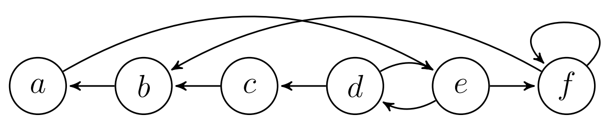

Fig. 4: Failure of **OU** removal persistence under semi-stable semantics.

Example 8 (Failure of **UO** *removal persistence under semi-stable semantics).* The AF shown in figure 2 has a unique semi-stable labelling $L = \{(a, \mathbf{U}), (b, \mathbf{I}), (c, \mathbf{O}), (d, \mathbf{U})\}$. If we remove the attack from a to c (labelled **U** and **O**) then L is no longer a semi-stable labelling. Instead we get the unique semi-stable labelling $\{(a, \mathbf{U}), (b, \mathbf{O}), (c, \mathbf{I}), (d, \mathbf{O})\}$.

Let us summarize the results obtained in this section. At the start we saw that under all semantics we consider, the **II**, **IU** and **UI** removal persistence are trivially satisfied. Furthermore, the **UU** and **IO** removal properties fail under all semantics, except for the stable semantics, which does not satisfy **IO** removal persistence but trivially satisfies all properties involving **U**-labelled arguments.

As for the remaining properties, it holds that the grounded, complete and preferred semantics are all similar in that they all satisfy **OO**, **OU**, **OI** and **UO** removal persistence. However, the semi-stable semantics satisfies none of these properties. Table 2 contains an overview of these results.

Grounded:

		X		
		O	**U**	**I**
Y	**O**	✓	✓	-
	U	✓	-	✓
	I	✓	✓	✓

Complete:

		X		
		O	**U**	**I**
Y	**O**	✓	✓	-
	U	✓	-	✓
	I	✓	✓	✓

Preferred:

		X		
		O	**U**	**I**
Y	**O**	✓	✓	-
	U	✓	-	✓
	I	✓	✓	✓

Stable:

		X		
		O	**U**	**I**
Y	**O**	✓	✓	-
	U	✓	✓	✓
	I	✓	✓	✓

Semi-Stable:

		X		
		O	**U**	**I**
Y	**O**	-	-	-
	U	-	-	✓
	I	-	✓	✓

Table 2: Overview of Removal Persistence Properties.

5 Skeptical Monotony

Suppose that two arguments x and y are labelled X and Y in all σ labellings of F. The XY addition persistence property then implies that all σ labellings of F are also σ labellings of F after adding an attack from x to y. In other words, XY addition persistence implies that no σ labelling gets destroyed. But is it also the case that no new labellings are created? This is the property that we consider in this section. We call it *skeptical XY monotony.*

Definition 10. *Let σ be a semanitcs and let $X, Y \in \{\mathbf{O}, \mathbf{I}, \mathbf{U}\}$. We say that σ satisfies XY skeptical monotony if and only if for all $(A, \rightsquigarrow) \in \mathcal{F}$ and $x, y \in A$: If for all $L \in \mathcal{L}_\sigma((A, \rightsquigarrow))$, $L(x) = X$ and $L(y) = Y$, then $\mathcal{L}_\sigma((A, \rightsquigarrow \cup \{(x, y)\})) \subseteq \mathcal{L}_\sigma((A, \rightsquigarrow))$.*

Like before, it is obvious that for some combinations of X and Y, the semantics we consider do not satisfy skeptical XY monotony. If x and y are both labelled **I** then adding an attack creates a complete labelling which is not a complete labelling of the initial AF. Thus, skeptical **II** monotony fails, and skeptical **IU** and **UI** monotony fail for the same reason, except under the stable semantics, which does not satisfy **II** skeptical monotony but trivially satisfies all properties involving **U** labelled arguments.

Proposition 7. *The grounded, complete, preferred and semi-stable semantics do not satisfy* **II**, **IU** *or* **UI** *skeptical monotony.*

Proposition 8. *The stable semantics satisfies* **UO**, **UU**, **UI**, **UO** *and* **UI** *skeptical monotony but does not satisfy* **II** *skeptical monotony.*

In the rest of this section we focus on the remaining properties, namely **OO**, **OU**, **OI**, **UO**, **UU** and **IO** skeptical monotony. Again, all these properties may be considered desirable. For example, **OO** skeptical monotony reflects the principle that adding an attack between two arguments that are both labelled **O** in every labelling does not lead to the creation of new points of view on argument acceptance. This is intuitive because the added attack does not introduce a conflict with respect to any of the labellings of the initial AF. The other skeptical monotony properties may be considered desirable for the same reason. We now check whether the semantics we consider satisfy these properties.

Let us start with the grounded semantics. Because the grounded labelling is unique, there is no difference between XY skeptical monotony and XY addition persistence. Thus, the following result follows immediately from theorem 1.

Theorem 8. *The grounded semantics satisfies* **OO***,* **OU***,* **UO***,* **UU** *and* **IO** *skeptical monotony but not* **OI** *skeptical monotony.*

We move on to the complete semantics.

Theorem 9. *The complete semantics satisfies* **OO***,* **OU***,* **UO** *and* **IO** *skeptical monotony but not* **UU** *or* **OI** *skeptical monotony.*

Proof. Let $(A, \rightsquigarrow) \in \mathcal{F}$ and $x, y \in A$. If $x \rightsquigarrow y$ we are done. In the remainder we assume that $x \not\rightsquigarrow y$. The **OO**, **OU**, **UO** and **IO** cases can be reduced to the following two cases:

- For all $L \in \mathcal{L}_{CO}((A, \rightsquigarrow))$, $L(y) = \mathbf{O}$. Then y is **O** in the grounded labelling of F. Hence there is a $z \in A$ s.t. $z \rightsquigarrow y$ and z is **I** in the grounded labelling of $(A, \rightsquigarrow)$. Furthemore since $x \not\rightsquigarrow y$ it holds that $x \neq z$. Theorem 1 implies that y is **O** and z is **I** in the grounded labelling of $(A, \rightsquigarrow \cup\{(x, y)\})$. Hence for all $L \in \mathcal{L}_{CO}((A, \rightsquigarrow \cup\{(x, y)\}))$, $L(y) = \mathbf{O}$ and $L(z) = \mathbf{I}$. Definition 6 implies that $\mathcal{L}_{CO}((A, \rightsquigarrow \cup\{(x, y)\})) \subseteq \mathcal{L}_{CO}((A, \rightsquigarrow))$.
- For all $L \in \mathcal{L}_{CO}((A, \rightsquigarrow))$, $L(x) = \mathbf{O}$ and $L(y) = \mathbf{U}$. Then x is **O** and y is **U** in the grounded labelling of F. Theorem 1 implies that x is **O** in the grounded labelling of $(A, \rightsquigarrow \cup\{(x, y)\})$. Hence for all $L \in \mathcal{L}_{CO}((A, \rightsquigarrow \cup\{(x, y)\}))$, $L(x) = \mathbf{O}$. Theorem 6 implies that $\mathcal{L}_{CO}((A, \rightsquigarrow \cup\{(x, y)\})) \subseteq \mathcal{L}_{CO}((A, \rightsquigarrow))$.

Counterexamples for **OI** and **UU** skeptical monotony are provided below.

Example 9 (Failure of **OI** *and* **UU** *skeptical monotony under the complete semantics).* Consider an AF with two arguments a and b where a attacks b. This AF has a unique complete labelling $\{(a, \mathbf{I}), (b, \mathbf{O})\}$. Adding an attack from b to a leads to an additional complete labelling $\{(a, \mathbf{O}), (b, \mathbf{I})\}$. A counterexample for **UU** skeptical monotony can be constructed by making a self-attacking.

The preferred semantics satisfies none of the remaining properties.

Theorem 10. *The preferred semantics does not satisfy* **OO***,* **OU***,* **OI***,* **UO***,* **UU** *or* **IO** *skeptical monotony.*

Example 9 is also a counterexample for **OI** and **UU** skeptical monotony under the preferred semantics. Counterexamples for the other cases are provided below.

Example 10 (Failure of **IO** *and* **OO** *skeptical monotony under preferred semantics).* The AF shown in figure 3 has one preferred labelling $\{(a, \mathbf{O}), (b, \mathbf{I}), (c, \mathbf{O}), (d, \mathbf{I}), (e, \mathbf{O}), (f, \mathbf{U})\}$. Adding an attack from e to c (both **O** in all preferred labellings) or from b to c (**I** and **O** in all preferred labellings) leads to a new preferred labelling that is not a preferred labelling of the initial AF: $\{(a, \mathbf{O}), (b, \mathbf{I}), (c, \mathbf{O}), (d, \mathbf{O}), (e, \mathbf{I}), (f, \mathbf{O})\}$.

Example 11 (Failure of **OU** *skeptical monotony under the preferred semantics).* The AF shown in figure 2 has one preferred labelling $\{(a, \mathbf{U}), (b, \mathbf{I}), (c, \mathbf{O}), (d, \mathbf{U})\}$. Adding add an attack from c to a (labelled **O** and **U** in all preferred labellings) leads to a new preferred labelling that is not a preferred labelling of the initial AF: $\{(a, \mathbf{O}), (b, \mathbf{O}), (c, \mathbf{I}), (d, \mathbf{O})\}$.

Example 12 (Failure of **UO** *skeptical monotony under the preferred semantics).* The AF shown in figure 1 has one preferred labelling $\{(a, \mathbf{I}), (b, \mathbf{O}), (c, \mathbf{I}), (d, \mathbf{O}), (e, \mathbf{U}), (f, \mathbf{U})\}$. Adding an attack from f to b (labelled **U** and **O** in all preferred labellings) leads to a new preferred labelling that is not a preferred labelling of the initial AF: $\{(a, \mathbf{I}), (b, \mathbf{O}), (c, \mathbf{O}), (d, \mathbf{I}), (e, \mathbf{O}), (f, \mathbf{I})\}$.

The stable semantics satisfies none of the XY skeptical monotony properties, except those that are trivially satisfied due to **U** labelled arguments, as established in proposition 8.

Theorem 11. *Stable semantics does not satisfy* **OO***,* **OI***,* **IO** *skeptical monotony.*

For the **OO** and **IO** case, example 10 can be turned into a counterexample by removing from the AF shown in figure 3 the argument f. For the **OI** case, example 9 counts as a counterexample.

Like the preferred semantics, the semi-stable semantics satisfies none of the remaining XY skeptical monotony properties.

Theorem 12. *The semi-stable semantics does not satisfy* **OO***,* **OU***,* **OI***,* **UO***,* **UU** *or* **IO** *skeptical monotony.*

All counterexamples for the preferred case (examples 9, 10, 11 and 12) also apply in the semi-stable case.

Let us summarize the results of this section. At the start we established that under all the semantics we consider, the **II**, **IU** and **UI** skeptical monotony properties fail. The exception is the stable semantics, which fails **II** skeptical monotony but trivially satisfies properties involving **U**-labelled arguments. Furthermore, the results for skeptical monotony under the grounded semantics coincide with the results of addition persistence. Finally, while the complete semantics still satisfies **OO**, **OU**, **UO** and **IO** skeptical monotony, the preferred and semi-stable semantics satisfy none of the skeptical monotony properties. Table 3 contains an overview of these results.

Grounded:

		X		
		O	**U**	**I**
Y	**O**	✓	✓	✓
	U	✓	✓	-
	I	-	-	-

Complete:

		X		
		O	**U**	**I**
Y	**O**	✓	✓	✓
	U	✓	-	-
	I	-	-	-

Preferred:

		X		
		O	**U**	**I**
Y	**O**	-	-	-
	U	-	-	-
	I	-	-	-

Stable:

		X		
		O	**U**	**I**
Y	**O**	-	✓	-
	U	✓	✓	✓
	I	-	✓	-

Semi-Stable:

		X		
		O	**U**	**I**
Y	**O**	-	-	-
	U	-	-	-
	I	-	-	-

Table 3: Overview of skeptical monotony properties.

6 Related Work

As we mentioned, our work extends the approach of Boella et al. [4, 5]. The refinement and abstraction principles that they studied are, in the single-extension case that they consider, equivalent to the addition and removal persistence properties. Indeed, the results we obtained for the grounded semantics coincide with theirs. As we discussed, our results extend theirs in a number of ways.

Our work is also related to earlier work we did on counterfactuals in argumentation. Sakama studied counterfactuals of the form $\alpha \Box\!\!\rightarrow \beta$ (meaning "if α were true then β would be true") where α and β are literals of the form $\mathbf{I}(x)$ or $\mathbf{O}(x)$ [15]. A counterfactual $\alpha \Box\!\!\rightarrow \beta$ is true w.r.t. an AF F and semantics σ if the change of F represented by α leads to the truth of β in all σ labellings of F. Here, the premise $\mathbf{O}(x)$ represents the addition of a new argument attacking x, while $\mathbf{I}(x)$ represents the removal of all attacks pointing to x. Rienstra's approach [14] is similar, and is based on a relation $\Vdash^F_\sigma$ determined by an AF F and semantics σ, between so called interventions and consequences. An intervention is a set of literals of the form $\mathbf{O}(x)$ or $\neg\mathbf{I}(x)$ that represent, respectively, addition of a new argument attacking x or of a new self-attacking argument attacking x. A formula ϕ is a consequence of an intervention Φ (written $\Phi \Vdash^F_\sigma \phi$) if the change represented by Φ leads to the truth of ϕ in all σ labellings of F. A number of properties that were studied using these models follow from the results that we have obtained. For example, failure of *Cautious Monotony* (if $\Phi \Vdash^F_\sigma \alpha$ and $\Phi \Vdash^F_\sigma \phi$ then $\Phi \cup \{\alpha\} \Vdash^F_\sigma \phi$) under the preferred semantics demonstrated by Rienstra follows from the failure of **IO** skeptical monotony (theorem 10). Similarly, the failure of *Cut* (if $\Phi \Vdash^F_\sigma \alpha$ and $\Phi \cup \{\alpha\} \Vdash^F_\sigma \phi$ then $\Phi \Vdash^F_\sigma \phi$) under the semi-stable semantics follows from the failure of **IO** addition persistence (theorem 5).

Cayrol et al. [9] studied the impact on the evaluation of an argumentation framework when new arguments and attacks are added. They define a number of properties to characterize this impact. Examples are changes leading to a larger, unique, or smaller set of extensions, changes that are monotonic (every extension of the old AF is included in an extension of the changed AF) and monotony w.r.t. an argument (every argument included in an extension of the old AF is also included in an extension of the changed AF). Then they study the relation between these properties and determine some conditions under which the addition of an argument leads to a certain type of change.

The results we obtained can be contrasted with the characterization of the *strong equivalence* relation between AFs [12]. Two AFs are strongly equivalent with respect to a semantics if they generate equivalent sets of extensions, and this equivalence is robust with respect to the addition of new arguments and attacks to both AFs. The characterization of this relation shows that the presence and absence of an attack from an argument x to an argument y are indistinguishable if certain syntactical conditions are met. Examples of such conditions are that one or both of x and y are self attacking, or that y attacks x. Oikarinen and Woltran [12] have determined the exact condition for a number of different semantics. The results that we obtained also imply that in certain cases, the presence and absence of an attack between two arguments x and y are indistinguishable. In our case, however, the conditions are not syntactic but depend on the status of x and y.

7 Conclusion and Future Work

We studied a number of properties concerning the behaviour of semantics for abstract argumentation when the AF changes. The properties are concerned with how the evaluation of an AF changes if an attack between two arguments is added or removed. The results provide insight into the behaviour of these semantics in a dynamic context. In particular, we have shown that the complete semantics satisfies all the intuitive properties that we have considered, that the grounded, preferred and stable semantics fail some of them, and that the semi-stable semantics fail all of them.

We plan to extend the current line of research in a number of ways. First of all, we plan to study weaker versions of the properties considered in this paper, look at skeptical monotony with respect to removal and obtain results with respect to semantics that were not considered here. Furthermore, we plan to study connections between the properties considered here and in other work on the behaviour of semantics of argumentation, such as strong equivalence [12], input/output behaviour [1] and directionality [2]. Finally, we expect that the results we have obtained will be useful in the ongoing research into modelling dynamical aspects of abstract argumentation, such as counterfactuals, abduction and revision.

References

1. Pietro Baroni, Guido Boella, Federico Cerutti, Massimiliano Giacomin, Leendert van der Torre, and Serena Villata. On the input/output behavior of argumentation frameworks. *Artificial Intelligence*, 217:144–197, 2014.
2. Pietro Baroni and Massimiliano Giacomin. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence*, 171(10-15):675–700, 2007.

3. Ringo Baumann and Gerhard Brewka. Expanding argumentation frameworks: Enforcing and monotonicity results. In *COMMA 2010, Desenzano del Garda, Italy, September 8-10, 2010*, volume 216 of *Frontiers in Artificial Intelligence and Applications*, pages 75–86. IOS Press, 2010.
4. Guido Boella, Souhila Kaci, and Leendert van der Torre. Dynamics in argumentation with single extensions: Abstraction principles and the grounded extension. volume 5590 of *LNCS*, pages 107–118. Springer, 2009.
5. Guido Boella, Souhila Kaci, and Leendert van der Torre. Dynamics in argumentation with single extensions: Attack refinement and the grounded extension (extended version). In *Argumentation in Multi-Agent Systems, 6th International Workshop, ArgMAS 2009, Budapest, Hungary, May 12, 2009. Revised Selected and Invited Papers*, volume 6057 of *LNCS*, pages 150–159. Springer, 2009.
6. Richard Booth, Souhila Kaci, Tjitze Rienstra, and Leendert van der Torre. A logical theory about dynamics in abstract argumentation. In *Scalable Uncertainty Management - 7th International Conference, SUM 2013, Washington, DC, USA, September 16-18, 2013. Proceedings*, volume 8078 of *LNCS*, pages 148–161. Springer, 2013.
7. Martin Caminada. On the issue of reinstatement in argumentation. In *Logics in Artificial Intelligence, 10th European Conference, JELIA 2006, Liverpool, UK, September 13-15, 2006, Proceedings*, volume 4160 of *LNCS*, pages 111–123. Springer, 2006.
8. Martin W. A. Caminada and Dov M. Gabbay. A logical account of formal argumentation. *Studia Logica*, 93(2-3):109–145, 2009.
9. Claudette Cayrol, Florence Dupin de Saint-Cyr, and Marie-Christine Lagasquie-Schiex. Change in abstract argumentation frameworks: Adding an argument. *Journal of Artificial Intelligence Research*, 38:49–84, 2010.
10. Sylvie Coste-Marquis, Sébastien Konieczny, Jean-Guy Mailly, and Pierre Marquis. On the revision of argumentation systems: Minimal change of arguments statuses. In *Principles of Knowledge Representation and Reasoning: The 14th International Conference, KR 2014, Vienna, Austria, July 20-24, 2014*. AAAI Press, 2014.
11. Phan Minh 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.
12. Emilia Oikarinen and Stefan Woltran. Characterizing strong equivalence for argumentation frameworks. *Artificial Intelligence*, 175(14-15):1985–2009, 2011.
13. John L. Pollock. *Cognitive Carpentry: A Blueprint for how to Build a Person*. MIT Press, Cambridge, MA, USA, 1995.
14. Tjitze Rienstra. *Argumentation in Flux - Modelling Change in the Theory of Argumentation*. PhD thesis, University of Luxembourg/University of Montpellier II, 2014.
15. Chiaki Sakama. Counterfactual reasoning in argumentation frameworks. In *Proceedings of COMMA 2014, Atholl Palace Hotel, Scottish Highlands, UK, September 9-12, 2014*, volume 266 of *Frontiers in Artificial Intelligence and Applications*, pages 385–396. IOS Press, 2014.