The relation between preferential model and argumentation semantics

Nico Roos
Department Knowledge Engineering, Maastricht University
P.O.Box 616, 6200 MD Maastricht, The Netherlands
email: roos@maastrichtuniversity.nl

Abstract

Although the preferential model semantics is the standard semantics for non-monotonic reasoning systems, it is not used for argumentation frameworks. For argumentation frameworks, instead, argumentation semantics are used. This paper studies the relation between the two types of semantics. Several argumentation semantics are related to additional constraints on the preference relation over states in the preferential model semantics. Moreover, based on the preferential model semantics a new argumentation semantics is proposed.

Introduction

Argumentation systems are becoming increasingly important for common sense and legal reasoning, negotiating agents, planning, and so on. An important issue is the underlying semantics of an argumentation system. The semantics of an argumentation system containing defeasible arguments is usually defined with respect to an *argumentation framework*. An argumentation framework is an abstraction of an argumentation system with respect to which an *argumentation semantics* is defined (Dung 1995).

Argumentation with defeasible arguments is a special case of *non-monotonic reasoning*. The *preferential model semantics* is the standard semantics for non-monotonic reasoning systems (Kraus, Lehmann, and Magidor 1990; Makinson 1988; 1994). This raises the question whether a preferential model semantics can be defined for argumentation frameworks? If a preferential model semantics can be defined, how does it relate to the well-known argumentation semantics? Finally, does it give us new insights with respect to how argumentation semantics should be defined?

Paper outline In the next section, the definitions of argumentation semantics and of preferential model semantics are given. Section "A preferential model semantics for argumentation frameworks" proposes a preferential model semantics for argumentation frameworks, and Section "Examples of preferential models" presents some examples of the proposed preferential model semantics. Section "The relation between the two types of semantics" addresses the relations between several argumentation semantics and the preferential model semantics. Section "Conclusion" concludes the paper.

Preliminaries

Argumentation semantics

We use Dung's argumentation framework as a starting point (Dung 1995).

Definition 1 *An argumentation framework is a couple $AF = \langle \mathcal{A}, \longrightarrow \rangle$ where $\mathcal{A}$ is a finite set of arguments and $\longrightarrow \subseteq \mathcal{A} \times \mathcal{A}$ is an attack relation over the arguments.*

For convenience, we extend the attack relation $\longrightarrow$ to sets of arguments.

Definition 2 *Let $A \in \mathcal{A}$ be an argument and let $\mathcal{S}, \mathcal{P} \subseteq \mathcal{A}$ be two sets of arguments. We define:*

- *$\mathcal{S} \longrightarrow A$ iff for some $B \in \mathcal{S}$, $B \longrightarrow A$.*
- *$A \longrightarrow \mathcal{S}$ iff for some $B \in \mathcal{S}$, $A \longrightarrow B$.*
- *$\mathcal{S} \longrightarrow \mathcal{P}$ iff for some $B \in \mathcal{S}$ and $C \in \mathcal{P}$, $B \longrightarrow C$.*

We wish to select coherent subsets of arguments $\mathcal{E}$ from the set of arguments $\mathcal{A}$ of the argumentation framework $AF = \langle \mathcal{A}, \longrightarrow \rangle$. Such a set of arguments $\mathcal{E}$ is called an *argument extension*. The arguments of an argument extension support propositions that give a coherent description of what might hold in the world. Clearly, a basic requirement of an argument extension is being *conflict-free*; i.e., no argument in an argument extension attacks another argument in the argument extension. Beside being conflict-free, we will use two notions of *defense against attacking arguments*. The notion of an *admissible set of arguments* and the notion of an *argument that is acceptable w.r.t. a set of arguments*. A set of admissible arguments defends itself against all attacking arguments and an argument that is acceptable w.r.t. a set of arguments, is defended by the set against all attacking arguments. In both cases, defense is realized by attacking the attacker.

Definition 3 *Let $AF = \langle \mathcal{A}, \longrightarrow \rangle$ be an argumentation framework and let $\mathcal{S} \subseteq \mathcal{A}$ be a set of arguments.*

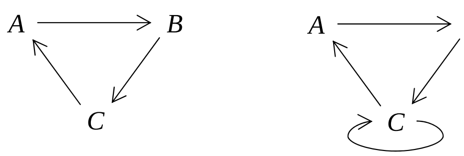
- *$\mathcal{S}$ is* conflict-free *iff $\mathcal{S} \not\longrightarrow \mathcal{S}$.*
- *$\mathcal{S}$ is admissible iff $\mathcal{S}$ is conflict-free and for every argument $A \in \mathcal{A}$: if $A \longrightarrow \mathcal{S}$, then $\mathcal{S} \longrightarrow A$.*
- *$A \in \mathcal{A}$ is acceptable w.r.t. $\mathcal{S}$ iff for every argument $B \in \mathcal{A}$, if $B \longrightarrow A$, then $\mathcal{S} \longrightarrow B$.*

Dung (1995) defines four argumentation semantics in terms of four types of argument extensions.[1]

Definition 4 *Let $AF = \langle \mathcal{A}, \longrightarrow \rangle$ be an argumentation framework and let $\mathcal{E} \subseteq \mathcal{A}$.*

- *$\mathcal{E}$ is a* stable extension *iff $\mathcal{E}$ is conflict-free and for every argument $A \in (\mathcal{A} - \mathcal{E})$, $\mathcal{E} \longrightarrow A$.*
- *$\mathcal{E}$ is a* preferred extension *iff $\mathcal{E}$ is maximal (w.r.t. $\subseteq$) admissible set of arguments.*
- *$\mathcal{E}$ is a* complete extension *iff (i) $\mathcal{E}$ is an admissible set of arguments, and (ii) every argument $A \in \mathcal{A}$ that is acceptable w.r.t. $\mathcal{E}$ belongs to $\mathcal{E}$.*
- *$\mathcal{E}$ is a* grounded extension *iff $\mathcal{E}$ is the minimal (w.r.t. $\subseteq$) complete extension.*

Note that the requirements of the stable semantics are quite strong. A stable extension defends itself against all arguments not belonging to the extension. As a result a stable extension need not exist. The odd loops of attacks shown in Figure 1 are examples of such problematic cases. The preferred, complete and grounded extension of each of the odd loops in Figure 1 is the empty set.



a **b**

Figure 1: Odd attack loops.

Baroni et al. (2005) argue that none of the extensions proposed by Dung handle odd loops adequately. They propose to handle odd loops in the same way as even loops by selecting conflict-free subsets of the arguments involved in the loops. Their *CF2 semantics* formalizes this point of view. In the example of Figure 1.a, it gives us the argument extensions: $\{A\}$, $\{B\}$ and $\{C\}$.

The CF2 semantics does not capture our intuitions with respect to the handling of odd loops in all circumstances. Consider for instance, the odd loop shown in Figure 1.b. According to the CF2 semantics, there are two argument extensions, namely $\{A\}$ and $\{B\}$. However, the self-attack of the argument C ensures that the argument A cannot depend on the argument B. Hence, the argument A should always be acceptable and the argument B never.

Justified arguments The existence of multiple argument extensions indicates uncertainty about which argument extension should describe the world. Using a skeptical view, we can only be certain of arguments that belong to every argument extension. These arguments are called the *justified* arguments. We can believe the propositions supported by the justified arguments because these arguments are present in every argument extension.

[1] In the last decade several new argumentation semantics have been proposed; for an overview, see (Baroni and Giacomin 2007; Bench-Capon and Dunne 2007).

Preferential model semantics

Preferential model semantics were introduced by Shoham (1987) and were subsequently extended to a general semantic theory by Makinson (1988) and Kraus et al. (1990). The definitions given below are based on the formalization given by Makinson in (1994).[2]

We start with a propositional language $\mathcal{L}$ for which we define the preferential model semantics. The preferential model semantics uses *preferential models* to define an agent's beliefs given its knowledge about the world.

Definition 5 *A preferential model $P = (S, \models, <)$ is a triple where:*

- *S is a set of states,*
- *$\models \subseteq (S \times \mathcal{L})$ is an arbitrary relation between states and propositions, called the* entailment relation[3],
- *$< \subseteq (S \times S)$ is an arbitrary relation between states, called the* preference relation.

Note that a preferential model does not specify what the states and the entailment relation exactly are. In general, one can view a state as an interpretation or a set of interpretations of propositional or first order logic. So, a state can be viewed as giving a (partial) description of the world. The entailment relation can then be viewed as a specification of the semantics of a proposition with respect to a state. For the moment, however, we do not consider such a restricted view on what the states and the entailment relation represent. Note that the preference relation denotes that we prefer a state s to a state s' if $s < s'$.[4]

A preferential model $P = (S, \models, <)$ can be used to specify that a state preferentially satisfies a proposition $\varphi \in \mathcal{L}$. Preferential entailment focusses on the preferred states among the states satisfying the proposition.

Definition 6 *Let $P = (S, \models, <)$ be a preferential model, $s \in S$ be a state and let $\varphi \in \mathcal{L}$ be a proposition.*

Then s preferentially satisfies *φ, denoted by $s \models_< \varphi$, iff $s \models \varphi$ and for no $s' \in S$: $s' < s$ and $s' \models \varphi$.*

We extend the notion of entailment of a proposition to set of propositions: $s \models \Sigma$ iff for every $\sigma \in \Sigma$, $s \models \sigma$. This immediately gives us the preferential entailment of a set of propositions: $s \models_< \Sigma$.[5] We need $s \models_< \Sigma$ to define the preferential consequences of a set of propositions Σ. Preferential consequences are those propositions that are entailed (satisfied) by all states that preferentially satisfy the set of propositions Σ. We will use the *preferential entailment operator $C_<$* to denote this set of consequences.

[2] The preferential model semantics should not be confused with the handling of conflicting arguments using a preference relation defined over the arguments; see for instance (Dimopoulos, Moraitis, and Amgoud 2008). A preference relation over arguments expresses in some way the strength of an argument while a preference relation over states expresses that the world should correspond to one of the preferred states.

[3] A state s is said to entail or satisfy a proposition φ iff $s \models \varphi$.

[4] For historical reasons, namely minimizing exceptions, preference is associated with minimality.

[5] $s \models_< \Sigma$ iff $s \models \Sigma$ and for no $s' < s$: $s' \models \Sigma$ iff $s \in min_< \|\Sigma\|$ where $\|\Sigma\| = \{s \in S \mid s \models \Sigma\}$.

Definition 7 *Let $P = (S, \models, <)$ be a preferential model, and let $\Sigma \subseteq \mathcal{L}$ be a set of propositions.*

The preferential entailment operator is defined as:

$$C_<(\Sigma) = \{\varphi \in \mathcal{L} \mid \text{ for all } s \in S, \text{ if } s \models_< \Sigma, \text{ then } s \models \varphi\}$$

A preferential model semantics for argumentation frameworks

In order to define a preferential semantics for an argumentation framework, we first have to determine how to interpret arguments. In an argumentation framework, an argument gives a warrant for some proposition. This warrant can be invalidated if it contains defeasible steps (Toulmin 1958). Therefore, in terms of a preferential model, *an argument expresses that we prefer states that satisfy the proposition supported by the argument to states that do not*.

In an argumentation framework, we have abstracted from the internal structure of an argument and from the specific proposition supported by the argument. Therefore, we cannot describe the preference expressed by an argument by preferring a state satisfying the proposition supported by the argument to a state that does not. However, instead of states satisfying propositions, we may consider *states satisfying arguments*. The idea is that a state gives a possible description of the world by specifying a set of valid arguments. Although an argumentation framework abstracts for the propositions supported by the arguments, we may view propositions supported by the valid arguments as describing what holds in a state.

The entailment relation $\models$ of a preferential model $P = (S, \models, <)$ must specify for each state the arguments that that we accept to give a correct description of what holds in world. The entailment relation is therefore a relation between states and arguments: $\models \subseteq (S \times \mathcal{A})$. So, the language $\mathcal{L}$ for which a preferential model is defined should consist of the set of arguments $\mathcal{A}$.

A state need not satisfy all arguments in $\mathcal{A}$. Especially, if an argument attacks another argument, a state cannot satisfy both arguments. A state cannot describe the world if it would satisfy arguments A and B while A attacks B ($A \longrightarrow B$).

Requirement 1 *Every state $s \in S$ must be conflict-free. That is, if $A \longrightarrow B$, then $s \models A$ and $s \models B$ may not hold at the same time.*

The set of arguments that are satisfied by a state $s \in S$ or a set of states $T \subseteq S$ will be denoted by $\mathcal{A}(s) = \{A \in \mathcal{A} \mid s \models A\}$ and $\mathcal{A}(T) = \bigcap_{s \in T} \mathcal{A}(s)$, respectively.

An attack relation between two arguments does not only express that both arguments cannot be entailed by one state. The attack relation also expresses a preference. *If an argument attacks another argument, we should prefer a state satisfying the attacking argument to a state satisfying the attacked argument*. Extending this preference generated by an attack relation between two arguments to the whole attack relation, we should prefer a state to another state if the arguments satisfied by the former state attack all arguments satisfied by the latter state but not by the former state.

Requirement 2 *An attack relation $\longrightarrow$ over arguments defines a preference relation over states.*

We prefer a state s to a state s' if and only if every argument satisfied by the state s' that is no longer satisfied by s is attacked by an argument in s.

Requirement 2 enables us to define a weak preference relation $\lesssim$ over a set of states S.

Definition 8 *Let $AF = \langle \mathcal{A}, \longrightarrow \rangle$ be an argumentation framework. Moreover, let S be a set of states and let $\models \subset (S \times \mathcal{A})$ be an entailment relations over states and arguments.*

The weak preference relation $\lesssim \subset (S \times S)$ is defined as:

$s \lesssim s'$ iff for every $B \in \mathcal{A}$ such that $s' \models B$ and $s \not\models B$, there is an $A \in \mathcal{A}$ such that $s \models A$ and $A \longrightarrow B$.

Note that the weak preference relation $\lesssim$ may contain cycles. Therefore, no minimum / preferred state may exist. The weak preference relation generated by the well known Nixon diamond contains such a cycle. States involved in a cycle have the indistinguishable preference.

States with the indistinguishable preference can be viewed as giving alternative descriptions of what holds in the world. These alternative descriptions should be made explicit. The way to do this is by deriving a strict preference relation $<$ from the weak preference relation $\lesssim$. This will guarantee that (*i*) a minimum state always exist, and (*ii*) multiple minimum states express indistinguishable preferences between these states.

We assume that $s \lesssim s'$ describes a strict preference if $s \lesssim s'$ is not part of a cycle. That is, if there does not exist a set of states $\{s_1, \ldots, s_n\}$ such that:

$$s \lesssim s' \lesssim s_1 \lesssim \ldots \lesssim s_n \lesssim s$$

Therefore, we define $s < s'$ as: $s \lesssim s'$ and $s' \not\lesssim^+ s$, where $\lesssim^+$ denotes the transitive closure of $\lesssim$.

Note that the preference relation $s \lesssim s'$ generated by the attack relation also holds if $\mathcal{A}(s') \subseteq \mathcal{A}(s)$. Nevertheless, we have to specify explicitly that we prefer states satisfying more (w.r.t. $\subset$) arguments. The reason is that a cycle in the preference relation $\lesssim$ may involve preferences such as $\mathcal{A}(s') \subset \mathcal{A}(s)$. If we have a cycle $s < s' \lesssim^+ s'' \lesssim s$ where $s < s'$ because $\mathcal{A}(s') \subset \mathcal{A}(s)$, we also have the cycle: $s' \lesssim^+ s'' \lesssim s'$. So, we have indistinguishable preference between both s'' and s, and s'' and s'. Since $\mathcal{A}(s') \subset \mathcal{A}(s)$, we should prefer s to s'.

Definition 9 *Let $AF = \langle \mathcal{A}, \longrightarrow \rangle$ be an argumentation framework, and let $\mathcal{L}$ be a language. Moreover, let $\lesssim$ be the weak preference relation generated by $\longrightarrow$.*

The preferential model $P = (S, \models, <)$ for the argumentation framework AF is defined as:

1. *S is a set of states and for every **conflict-free** set of arguments $\mathcal{S} \subseteq \mathcal{A}$, there is* exactly *one state $s \in S$;*
2. *$\models \subseteq (S \times \mathcal{A})$ where for every conflict-free set of arguments $\mathcal{S} \subseteq \mathcal{A}$, there is state $s \in S$ such that $\mathcal{A}(s) = \mathcal{S}$;*
3. *$s < s'$ iff $\mathcal{A}(s') \subset \mathcal{A}(s)$, or $s \lesssim s'$ and $s' \not\lesssim^+ s$.*

The first item in the above definition states that for every conflict-free set of arguments there is a state. The second item states that the entailment relation is defined between states and arguments (the language). Moreover, it specifies that for every consistent set of arguments there is a state satisfying exactly these arguments. The third item specifies the preference relation using the weak preference relation.

Examples of preferential models

Before formally analyzing the relation between the above defined preferential model semantics and the argumentation semantics, we will first look at some examples.

The first example is an argumentation framework with three arguments A, B and C and the attack relation shown in Figure 2. Figure 3.a shows the transitive reduction[6] of

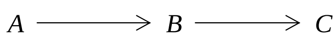
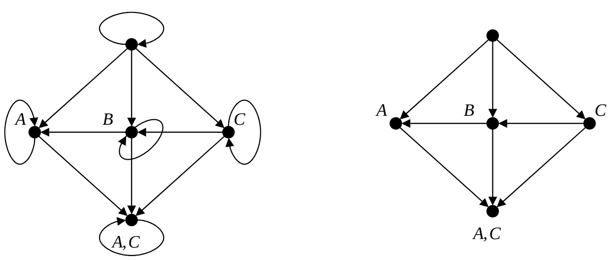


Figure 2: An attack chain.

the preference relation $\lesssim$ generated by the argumentation framework. Figure 3.b shows the transitive reduction of the preference relation $<$. In this figure we see that there is one minimum state, which corresponds with the argument extension of the grounded, the preferred, the stable and the CF2 semantics.



a **b**

Figure 3: Preferences generated by an attack chain.

The second example is an argumentation framework with two arguments A and B and the attack relation shown in Figure 4. Figure 5.a shows the preference relation $\lesssim$ generated

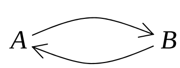
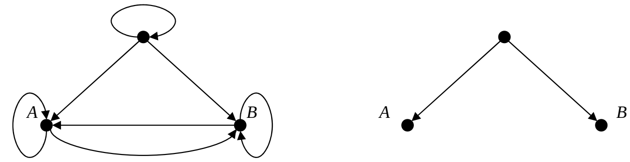


Figure 4: An even attack loop.

by the argumentation framework. Figure 5.b shows the preference relation $<$. In this figure we see that there are two minimum states, which correspond with the two argument extensions of the preferred, the stable and the CF2 semantics. The argument extension of the grounded semantics is consistent with both preferred states.

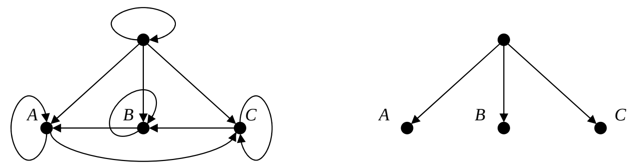
The third example is an argumentation framework with three arguments A, B and C and the attack relation shown in Figure 1.a, which forms an odd loop. Figure 6.a shows the preference relation $\lesssim$ generated by the argumentation framework. Figure 6.b shows the preference relation $<$. In this figure we see that there are three minimum states, which

[6]The transitive reduction of $\lesssim$ is the minimal sub-relation R of $\lesssim$ such that $\lesssim$ is contained in the transitive closure of R: $\lesssim \subseteq R^+$.



a **b**

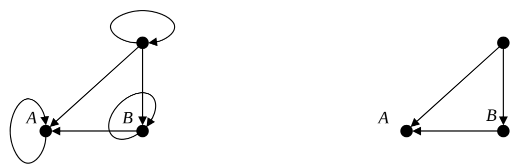
Figure 5: Preferences generated by an even attack loop.



a **b**

Figure 6: Preferences generated by an odd attack loop.

correspond with the three argument extensions of the CF2 semantics. The argument extension of the grounded and preferred semantics are consistent with the preferred states.



a **b**

Figure 7: Preferences generated by an odd attack loop with self-attack.

The fourth example is also an argumentation framework with three arguments A, B and C and the attack relation between them forming an odd loop. However, as shown in Figure 1.b, argument C also attacks itself. Figure 7.a shows the preference relation $\lesssim$ generated by the argumentation framework. Figure 7.b shows the preference relation $<$. In this figure we see that, unlike the previous example, here there is one minimum state. The extension of the grounded and preferred semantics are consistent with this preferred state. The CF2 semantics specifies two argument extensions for the argumentation framework: $\{A\}$ and $\{B\}$. Only the first CF2-extension corresponds with the minimum state.

The relation between the two types of semantics

The examples presented in the previous section suggest that there is a relation between the proposed preferential model semantics and some of the well known argumentation semantics. In this section we will investigate this relation. Moreover, we define a new argumentation semantics and show that it is equivalent to the preferential model semantics. We conclude the section by showing that the closure property *cumulativity* holds for the preferential model semantics and therefore also for the new argumentation semantics.

Preferred states and conflict-free set of arguments Given a preferential model, we are interested in the minimum / preferred states. The first thing that we can observe is that such a minimum state satisfies a maximal conflict-free set of arguments.

Proposition 1 *Let $P = (S, \models, <)$ be a preferential model for an argumentation framework $AF = \langle \mathcal{A}, \longrightarrow \rangle$.*

For every minimum s in S (w.r.t. $<$): $\mathcal{A}(s) = \{A \in \mathcal{A} \mid s \models A\}$ is a maximal conflict-free set of arguments.

The relation with argumentation semantics The above proposition suggests a relation between argument extensions and preferred / minimum states of the preferential model. The following theorems make this relation explicit. The first theorem establishes a relation between the preferential model semantics and the *stable semantics*. A stable extension defends itself against all arguments not belonging to the extension. This implies that an argument A that is not attacked by the stable extension $\mathcal{E}$ should belong to $\mathcal{E}$. A cannot attack $\mathcal{E}$ because then the stable extension cannot exist. Therefore, a state s representing the stable extension $\mathcal{E}$ must be a preferred / minimum (w.r.t. $<$) state. Moreover, arguments attacking the stable extension $\mathcal{E}$ result in weakly preferring a state s' to s. Since a stable extension defends itself against all arguments that do not belong to the extension, s should be weakly preferred to s'.

Theorem 1 *Let $P = (S, \models, <)$ be a preferential model for an argumentation framework $AF = \langle \mathcal{A}, \longrightarrow \rangle$.*

$\mathcal{A}(s)$ is a stable extension *iff s is a minimum state in S, and for any state $s' \in S$, if $s' \lesssim s$, then $s \lesssim s'$.*

The restriction on the weak preference relation that has been used to establish a relation with the stable semantics can also be used to establish a relation with the *preferred semantics*. The state s representing a preferred extension should also defend itself against all attacking arguments. However, the state s need not be a preferred / minimum state. In fact, the state representing a preferred extension is a minimum state among the states defending themselves against all attacking arguments. To identify the states defending themselves against all attacking arguments, we should only consider weakly preferred states where the preference is completely due to attacking arguments. The following definition formalizes the restriction on the weak preference relation and defines the states that defend themselves against all attacking arguments. The latter states are called admissible states.

Definition 10 *Let $\lesssim$ be a preference relation as defined in Definition 8.*

The preference relation that is the result of attacking arguments only is defined as:

$$\underset{\sim}{\ll} = \{(s, s') \mid s \lesssim s', \forall A \in (\mathcal{A}(s) - \mathcal{A}(s')) \colon A \longrightarrow \mathcal{A}(s')\}$$

$s \in S$ is an admissible state *iff for every state $s' \in S$ such that $s' \underset{\sim}{\ll} s$, $s \lesssim s'$.*

A preferred / minimum (w.r.t. $<$) state among the admissible states correspond to a preferred extension.

Theorem 2 *Let $P = (S, \models, <)$ be a preferential model for an argumentation framework $AF = \langle \mathcal{A}, \longrightarrow \rangle$.*

$\mathcal{A}(s)$ is a preferred extension *iff s is a minimum (w.r.t. $<$) admissible state in S.*

An argument is acceptable with respect to a set of arguments if the latter defends the former against all attacking arguments. We can define a somewhat similar notion in terms of preference over states of a preferential model. We introduce the notion of a state s that is acceptable with respect to another state s'. Since states are conflict-free, we will make use of the property that an argument can be added to a conflict-free set of arguments without introducing conflicts if the argument is acceptable w.r.t. this set. We therefore require that the state s satisfies at least the same set of arguments as the state s'. We must also ensure that arguments satisfied by s' defend the arguments satisfied by s against all attacking arguments. Arguments attacking the arguments of s can be described by states s'' such that $s'' \underset{\sim}{\ll} s$, and the defense by $s' \lesssim s''$.

Definition 11 *A state s* is acceptable with respect to *a state s' iff $\mathcal{A}(s') \subseteq \mathcal{A}(s)$ and for every state $s'' \in S$ if $s'' \underset{\sim}{\ll} s$, then $s' \lesssim s''$.*

We can now establish the relation with complete semantics.

Theorem 3 *Let $P = (S, \models, <)$ be a preferential model for an argumentation framework $AF = \langle \mathcal{A}, \longrightarrow \rangle$.*

$\mathcal{A}(s)$ is a complete extension *iff s is an admissible state in S and s is the only state that is acceptable with respect to s.*

The grounded semantics selects the unique subset minimal complete extension.

Theorem 4 *Let $P = (S, \models, <)$ be a preferential model for an argumentation framework $AF = \langle \mathcal{A}, \longrightarrow \rangle$.*

$\mathcal{A}(s)$ is a grounded argument extension *iff s is a maximum (w.r.t. $<$) state among the states in S that are both admissible and for which s is the only state acceptable with respect to s.*

The above four theorems imply that the set of preferred conclusions $C_<(\varnothing)$ of the preferential model correspond with the set of justified arguments.

Corollary 1 *Let $P = (S, \models, <)$ be a preferential model for an argumentation framework $AF = \langle \mathcal{A}, \longrightarrow \rangle$.*

$C_<(\varnothing)$ is the set of justified arguments of the stable, preferred, complete and grounded semantics if the restrictions of Theorems 1, 2, 3 and 4 are applied, respectively.

A new argumentation semantics The preferential model semantics can be used to define a new argumentation semantics. The idea is to use the preference relation on states to give a new definition of acceptable arguments. We first define a preference relation on sets of arguments.

Definition 12 *Let $AF = \langle \mathcal{A}, \longrightarrow \rangle$ be an argumentation framework and let $\mathcal{S}, \mathcal{T} \subseteq \mathcal{A}$ be two conflict-free sets of arguments.*

The set of arguments $\mathcal{T}$ is at least as acceptable as *$\mathcal{S}$, denoted by $\mathcal{T} \succsim \mathcal{S}$, iff for every argument $B \in \mathcal{S} - \mathcal{T}$ there is an argument $A \in \mathcal{T} - \mathcal{S}$ such that: $A \longrightarrow B$.*

The above defined $\succsim$-relation may contain cycles. Therefore, to identify a maximally acceptable set of arguments, similar to the definition of the preference relation $<$ of a preferential model (Definition 9), we have to take into account loops of preferences.

Definition 13 *Let $AF = \langle \mathcal{A}, \longrightarrow \rangle$ be an argumentation framework and let $\mathcal{T} \succsim \mathcal{S}$ be an acceptability relation.*

A conflict-free set of arguments $\mathcal{E} \subseteq \mathcal{A}$ is a pm-extension *iff for every conflict-free set of arguments $\mathcal{T} \subseteq \mathcal{A}$, if $\mathcal{T} \succsim \mathcal{E}$, then $\mathcal{E} \succsim^{+} \mathcal{T}$ and $\mathcal{E} \not\subset \mathcal{T}$.*

Because the pm-extensions are based on the preferential model semantics, it is not difficult to show that the minimum states correspond with the pm-extensions.

Theorem 5 *Let $AF = \langle \mathcal{A}, \longrightarrow \rangle$ be an argumentation framework and let $P = (S, \models, <)$ be a corresponding preferential model.*

For every pm-extension $\mathcal{E} \subseteq \mathcal{A}$, there is a minimum state $s \in S$ such that $\mathcal{E} = \mathcal{A}(s)$, and vice versa.

The closure property Cumulativity is generally considered to be a desirable property of non-monotonic reasoning systems. For argumentation frameworks it is less important since we normally determine the justified arguments starting form an empty set of arguments; i.e., $C_<(\varnothing)$. It may, however, be useful in creating a proof theory for the proposed preferential model semantics.

Proposition 2 *Let $P = (S, \models, <)$ be a preferential model for an argumentation framework $AF = \langle \mathcal{A}, \longrightarrow \rangle$. Moreover, let the attack relation $\longrightarrow$ contain a finite number of elements.*

Then the consequence operator $C_<(\cdot)$ defined by the preferential model $P = (S, \models, <)$, is cumulative*:*

$$\textit{if } \Sigma \subseteq \Gamma \subseteq C_<(\Sigma), \textit{ then } C_<(\Sigma) = C_<(\Gamma)$$

Conclusion

In this paper a preferential model semantics for argumentation frameworks is proposed and is compared with several well known argumentation semantics.

Argumentation semantics are based on the notion of defense against attacking arguments. The way this notion of defense is applied leads to different semantics. In the presences of odd loops, the adequacy of the notion has been debated.

The proposed preferential model semantics is based on a different notion; preference over states. Here, states are viewed as descriptions of the world that specify sets of valid arguments. The attack relation over arguments is used to define a preference relation over states. A state in which the attacking argument is valid is preferred to a state in which the attacked argument is valid.

In the absence of odd loops in the argumentation framework, the preferential model semantics supports the same argument extensions as the stable semantics. In general, the argument extensions supported by the different argumentation semantics proposed by Dung correspond to different restrictions on the preference relation of the preferential model semantics. In the presence of odd loops, without restrictions on the preference relation, the preferential model semantics support a different set of argument extensions. In the absence of nested odd loops, the set of supported argument extension correspond to those supported by the CF2 semantics. The way the preferential model semantics handles odd loops, especially, nested odd loops, seems to be more in line with our intuitions. Based on the preferential model semantics, a new equivalent argumentation semantics has been defined.

Appendix

Proof of Proposition 1 Suppose that $\mathcal{A}(s)$ is not a maximal conflict-free set of argument. Then there is a conflict-free set of arguments $\mathcal{S} \subseteq \mathcal{A}$ such that $\mathcal{A}(s) \subset \mathcal{S}$. According to item 2 of Definition 9, there is a state $s' \in S$ such that $\mathcal{S} = \mathcal{A}(s')$. Then, according to item 3 of Definition 9, $s' < s$. Contradiction. □

Proof of Theorem 1 (⇒) Let $\mathcal{A}(s)$ be a stable argument extension. Since every argument not in $\mathcal{A}(s)$ is attacked by $\mathcal{A}(s)$, $\mathcal{A}(s)$ is a maximal conflict-free set of arguments. Therefore, for no $s' \in S$, $\mathcal{A}(s) \subset \mathcal{A}(s')$.

Suppose that there is $s' \in S$ such that $s' < s$. Since $\mathcal{A}(s) \not\subset \mathcal{A}(s')$, according to item 3 of Definition 9 and Definition 8, there is an argument $A \in \mathcal{A} - \mathcal{S}$ such that $A \longrightarrow \mathcal{S}$. Since $\mathcal{S}$ is a stable arguments extension, for every argument $A \in \mathcal{A} - \mathcal{S}$, $\mathcal{S} \longrightarrow A$. Therefore, $s \lesssim s'$. Moreover, since $\mathcal{A}(s) \not\subset \mathcal{A}(s')$, $s' \not< s$. Contradiction.

Hence, s is minimum / preferred state in S such that $\mathcal{A}(s)$.

(⇐) Let s be a minimum state in S, and for any state $s' \in S$, if $s' \lesssim s$, then $s \lesssim s'$. According to Proposition 1, $\mathcal{A}(s)$ is a maximal conflict-free set of arguments.

Suppose that $\mathcal{A}(s)$ is not a stable argument extension. Then for some $A \in (\mathcal{A} - \mathcal{A}(s))$, $\mathcal{A}(s) \not\longrightarrow A$. Since $\mathcal{A}(s)$ is maximal conflict-free set, $A \longrightarrow \mathcal{A}(s)$. Let $\mathcal{A}(s') = \{A\} \cup (\mathcal{A}(s) - \{B \mid A \longrightarrow B\})$. According to Definition 8, $s' \lesssim s$. This implies $s \lesssim s'$. Therefore, there must be an argument $B \in \mathcal{A}(s)$ such that $B \longrightarrow A$. Contradiction.

Hence, $\mathcal{A}(s)$ is not a stable argument extension. □

Lemma 1 *Let $P = (S, \models, <)$ be a preferential model for an argumentation framework $AF = \langle \mathcal{A}, \longrightarrow \rangle$.*

$\mathcal{A}(s)$ is admissible iff s is an admissible state.

Proof (⇒) Let $\mathcal{A}(s)$ be admissible.

Suppose that $s' \stackrel{\ll}{\sim} s$. Then for every $B \in (\mathcal{A}(s') - \mathcal{A}(s))$, $B \longrightarrow \mathcal{A}(s)$. Since $\mathcal{A}(s)$ is admissible, $\mathcal{A}(s) \longrightarrow B$. Therefore, $s \lesssim s'$.

Hence, for every $s' \stackrel{\ll}{\sim} s$ implies $s \lesssim s'$ and therefore, s is an admissible state.

(⇐) Let $s \in S$ be an admissible state. Then, $s' \stackrel{\ll}{\sim} s$ imply $s \lesssim s'$ for every $s' \in S$.

Suppose that $\mathcal{A}(s)$ is not admissible. Then there must be an argument B such that $B \longrightarrow \mathcal{A}(s)$ and $\mathcal{A}(s) \not\longrightarrow B$. Let $\mathcal{A}(s') = \{B\} \cup (\mathcal{A}(s) - \{C \mid B \longrightarrow C\})$. Clearly, $s' \stackrel{\ll}{\sim} s$. Since $\mathcal{A}(s) \not\longrightarrow B$, $s \not\lesssim s'$. Contradiction.

Hence, $\mathcal{A}(s)$ is admissible. □

Proof of Theorem 2 $\mathcal{A}(s)$ is a *preferred extension* iff $\mathcal{A}(s)$ is a maximal, w.r.t. $\subseteq$) admissible set. According to Lemma 1, $\mathcal{A}(s')$ is admissible iff for every $s'' \in S$: $s'' \underset{\sim}{\ll} s'$ implies $s' \lesssim s''$.

Therefore, $\mathcal{A}(s)$ is a *preferred argument extension* iff s maximizes (w.r.t. $\subseteq$) $\mathcal{A}(s)$ among the states $\{s' \in S \mid \forall s'' \in S$: $s'' \underset{\sim}{\ll} s'$ implies $s' \lesssim s''\}$. Hence, $\mathcal{A}(s)$ is a *preferred argument extension* iff s is a minimum state w.r.t. $<$ in $\{s' \in S \mid \forall s'' \in S$: $s'' \underset{\sim}{\ll} s'$ implies $s' \lesssim s''\}$; i.e., s is a minimum, (w.r.t. $<$) admissible state in S. □

Lemma 2 *Let $P = (S, \models, <)$ be a preferential model for an argumentation framework $AF = \langle \mathcal{A}, \longrightarrow \rangle$. Moreover, let $\mathcal{A}(s)$ be an admissible set of arguments.*

Every argument acceptable w.r.t. $\mathcal{A}(s)$ belongs to $\mathcal{A}(s)$ iff s is the only state that is acceptable with respect to s.

Proof ($\Rightarrow$) Let very argument A acceptable w.r.t. $\mathcal{A}(s)$ belongs to $\mathcal{A}(s)$.

Suppose s' is acceptable w.r.t. s and $s \neq s'$. Let $s'' \in S$ be a state such that $s'' \underset{\sim}{\ll} s'$. Then for every $B \in (\mathcal{A}(s'') - \mathcal{A}(s'))$, $B \longrightarrow \mathcal{A}(s')$. Since s' is acceptable w.r.t. s, $s \lesssim s''$. Therefore, $\mathcal{A}(s) \longrightarrow B$. Hence, for every $A \in (\mathcal{A}(s') - \mathcal{A}(s))$, A is acceptable w.r.t. $\mathcal{A}(s)$. But then, $A \in \mathcal{A}(s)$, contradicting $s \neq s'$.

Hence, s is the only state that is acceptable with respect to s.

($\Leftarrow$) Let s be the only state that is acceptable with respect to s.

Suppose that an argument A acceptable w.r.t. $\mathcal{A}(s)$ does not belong to $\mathcal{A}(s)$. Since A is acceptable w.r.t. $\mathcal{A}(s)$ and $\mathcal{A}(s)$ is admissible, there is a state $s' \in S$ such that $\mathcal{A}(s') = \{A\} \cup \mathcal{A}(s)$.

Suppose $s'' \underset{\sim}{\ll} s'$. Then for every $B \in (\mathcal{A}(s'') - \mathcal{A}(s'))$, $B \longrightarrow \mathcal{A}(s')$. Since A is an acceptable w.r.t. $\mathcal{A}(s)$, $\mathcal{A}(s) \longrightarrow B$. Therefore, $s \lesssim s''$.

Hence, s' is acceptable w.r.t. s and $s \neq s'$. Contradiction.

Hence, every argument acceptable w.r.t. $\mathcal{A}(s)$ belongs to $\mathcal{A}(s)$. □

Proof of Theorem 3 $\mathcal{A}(s)$ is a *complete extension* iff $\mathcal{A}(s)$ is admissible and every argument acceptable w.r.t. $\mathcal{A}(s')$ belongs to $\mathcal{A}(s')$. According to Lemma 1, $\mathcal{A}(s)$ is admissible iff s is an admissible state. According to Lemma 2, every argument acceptable w.r.t. $\mathcal{A}(s)$ belongs to $\mathcal{A}(s)$ iff s is the only state that is acceptable with respect to s. Therefore, $\mathcal{A}(s)$ is a *complete extension* iff s is an admissible state in S and s is the only state that is acceptable with respect to s. □

Proof of Theorem 4 $\mathcal{A}(s)$ is a *grounded extension* iff $\mathcal{A}(s)$ in a minimal, w.r.t. $\subseteq$, complete extension. Therefore, s is a maximum (w.r.t. $<$) state among the states in S that are both admissible and for which s is the only state acceptable with respect to s. □

Lemma 3 *$\mathcal{A}(s)$ is more acceptable than $\mathcal{A}(s')$, $\mathcal{A}(s) \gtrsim \mathcal{A}(s')$, iff $s \lesssim s'$.*

Proof ($\Rightarrow$) Let $\mathcal{A}(s) \gtrsim \mathcal{A}(s')$. Then, by Definition 12, for every argument $B \in \mathcal{A}(s') - \mathcal{A}(s)$ there is an argument $A \in \mathcal{A}(s) - \mathcal{A}(s')$ such that $A \longrightarrow B$. Therefore, for every argument B such that $s' \models B$ and $s \not\models B$ there is an argument A such that $s \models A$, $s' \not\models A$ and $A \longrightarrow B$. Hence, by Definition 9, $s < s'$.

($\Leftarrow$) Let $s < s'$. Then, by Definition 9, for every argument B such that $s' \models B$ and $s \not\models B$ there is an argument A such that $s \models A$ and $A \longrightarrow B$. B such that $s' \models B$ and $s \not\models B$ implies $B \in \mathcal{A}(s') - \mathcal{A}(s)$. Moreover, $s \models A$ implies $A \in \mathcal{A}(s)$, and $A \longrightarrow B$ together with $s' \models B$ implies $A \not\in \mathcal{A}(s')$ (conflict-free). Hence, by Definition 13, $\mathcal{A}(s) \gtrsim \mathcal{A}(s')$. □

Proof of Theorem 5 ($\Rightarrow$) Let $\mathcal{S}$ be a pm-extension. By Definition 13, $\mathcal{S}$ is conflict-free. Therefore there is an $s \in S$ such that $\mathcal{A}(s) = \mathcal{S}$.

Suppose s is not a minimum state. Then there is a $s' \in S$ such that $s' < s$. According to item 4 of Definition 9, $\mathcal{A}(s) \subset \mathcal{A}(s')$, or $s' \lesssim s$ and $s \not\lesssim s'$. According to Definition 13, $\mathcal{A}(s) \not\subset \mathcal{A}(s')$. Therefore, $s' \lesssim s$ and $s \not\lesssim s'$. This is equivalent to $\mathcal{A}(s) \gtrsim \mathcal{A}(s')$ and $\mathcal{A}(s) \not\gtrsim^{+} \mathcal{A}(s')$ according to Lemma 3. Hence according to Definition 13, $\mathcal{S} = \mathcal{A}(s)$ is not a pm-extension. Contradiction.

Hence, s is minimum state.

($\Leftarrow$) Let s be a minimum state. Then $\mathcal{A}(s)$ is a conflict-free set of arguments.

Suppose that $\mathcal{A}(s)$ is not a pm-extension. Then there is a conflict-free set of arguments $\mathcal{T} \subseteq \mathcal{A}$ such that $\mathcal{A}(s) \subset \mathcal{T}$, or $\mathcal{T} \gtrsim \mathcal{A}(s)$ and $\mathcal{A}(s) \gtrsim \mathcal{T}$. Since $\mathcal{T}$ is conflict-free, there is a state $s' \in S$ such that $\mathcal{A}(s) = \mathcal{T}$. According to Lemma 3, $\mathcal{A}(s) \subset \mathcal{A}(s')$, or $s' \lesssim s$ and $s \not\lesssim^{+} s'$. Then according to item 4 of Definition 9, there is an $s' \in S$ such that $s' < s$. Hence, s is not a minimum state. Contradiction.

Hence, $\mathcal{A}(s)$ is a pm-extension. □

The proof of Proposition 2 is omitted due to lack of space.

References

Baroni, P., and Giacomin, M. 2007. On principle-based evaluation of extension-based argumentation semantics. *Artificial Intelligence* 171:675–700.

Baroni, P.; Giacomin, M.; and Guida, G. 2005. Scc-recursiveness: a general schema for argumentation semantics. *Artificial Intelligence* 168:162–210.

Bench-Capon, T., and Dunne, P. E. 2007. Argumentation in artificial intelligence. *Artificial Intelligence* 171:619–641.

Dimopoulos, Y.; Moraitis, P.; and Amgoud, L. 2008. Theoretical and computational properties of preference-based argumentation. In *ECAI 2008*, 463–467.

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

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

Makinson, D. 1988. General theory of cumulative inference. In Reinfrank, M.; de Kleer, J.; Ginsberg, M. L.; and Sandewall, E., eds., *Proceedings of the second International Workshop on Non-monotonic Reasoning*, 1–18. Springer-Verlag.

Makinson, D. 1994. Nonmonotonic reasoning and uncertain reasoning. In Gabbay, D., ed., *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 3. 35–110.

Shoham, Y. 1987. A semantical approach to nonmonotonic logics. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, 388–392.

Toulmin, S. 1958. *The uses of argument*. Cambridge University Press.