Extending Argumentation to make good Decisions

Yannis Dimopoulos, Pavlos Moraïtis, Leila Amgoud

Extending Argumentation to make good Decisions

Yannis Dimopoulos[1], Pavlos Moraitis[2], and Leila Amgoud[3]

[1] Dept. of Computer Science, University of Cyprus, Nicosia, Cyprus
yannis@cs.ucy.ac.cy
[2] Laboratory of Informatics Paris Descartes (LIPADE), Paris Descartes University, Paris, France
pavlos@mi.parisdescartes.fr
[3] Paul Sabatier University, Toulouse, France
amgoud@irit.fr

Abstract. Argumentation has been acknowledged as a powerful mechanism for automated decision making. In this context several recent works have studied the problem of accommodating preference information in argumentation. The majority of these studies rely on Dung's abstract argumentation framework and its underlying acceptability semantics.
In this paper we show that Dung's acceptability semantics, when applied to a preference-based argumentation framework for decision making purposes, may lead to counter intuitive results, as it does not take appropriately into account the preference information. To remedy this we propose a new acceptability semantics, called *super-stable extension* semantics, and present some of its properties. Moreover, we show that argumentation can be understood as a multiple criteria decision problem, making in this way results from decision theory applicable to argumentation.

1 Introduction

In many decision making situations we are confronted with a set of alternatives or options each of which has its own advantages that can be expressed as different arguments supporting that alternative. For instance, in a car purchase scenario, one argument that supports small cars is that they have low running cost, while another argument that favors big cars is that they have better safety features. The final decision is usually based upon the preferences one has over the arguments, or more generally how arguments relate to each other.

It is therefore not surprising that during the last years, argumentation has been acknowledged as a powerful mechanism for automating the decision making process of autonomous agents. Several recent works (see e.g. [1–5]) have emphasized the role of agents' preferences in the evaluation of their arguments within a particular class of argumentation frameworks called *preference-based argumentation* frameworks. The majority of these frameworks are using the acceptability semantics of the Dung's abstract argumentation framework [6].

In [7], it has been shown that preference-based argumentation under the stable extension semantics is essentially a method for making decisions that

are supported by "good" or "strong" arguments. Roughly speaking, a set of arguments E is a stable extension if every argument of E is strictly preferred to any other argument that is not included in E.

In this work we show that the stable extensions semantics of Dung's framework when applied to decision making may lead to counterintuitive results and therefore fail to deliver the correct conclusions.

More precisely, we show that stable extensions consider as equally good two sets of arguments (and therefore the options they support), although for every argument of the second set, the first set contains a more preferred argument. One may understand that in this case the agent could randomly select an option that is supported either by an argument from the first or the second set and this could be a wrong decision if these arguments support incorrect conclusions. This problem relates to a similar problem identified independently by Horty in [8] in the context of the use of argumentation for defeasible reasoning. For this reason we propose a new semantics called *super-stable* extension which allows to fix this problem.

Finally, in this paper we show the correspondence between argumentation and multi-criteria decision making. Then we emphasize that an aggregation method like *regime* [9] can be an alternative approach for defining a ranking on the set of arguments supporting the options and consequently on the options themselves.

2 Basics of argumentation

Argumentation is a reasoning model based on the following main steps: i) constructing *arguments* and counter-arguments, ii) defining the *strengths* of those arguments, and iii) defining the *justified conclusions*. Argumentation systems are built around an underlying logical language and an associated notion of logical consequence, defining the notion of argument. The argument construction is a monotonic process: new knowledge cannot rule out an argument but only gives rise to new arguments which may interact with the first argument. Arguments may be conflicting for different reasons.

Definition 1 (Argumentation system [6]) *An* argumentation system *is a pair* $T = (\mathcal{A}, \mathcal{R})$. $\mathcal{A}$ *is a set of arguments and* $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$ *is an attack relation. We say that an argument* a *attacks an argument* b *iff* $(a, b) \in \mathcal{R}$.

Among all the arguments, it is important to know which arguments to keep for inferring conclusions. In [6], different acceptability semantics have been proposed. The basic idea behind these semantics is the following: for a rational agent, an argument a_i is acceptable if he can defend a_i against all attacks. All the arguments acceptable for a rational agent will be gathered in a so-called *extension*. An extension must satisfy a consistency requirement and must defend all its elements.

Definition 2 (Conflict-free, Defence [6]) *Let* $\mathcal{B} \subseteq \mathcal{A}$, *and* $a_i \in \mathcal{A}$.

– $\mathcal{B}$ *is* conflict-free *iff* $\nexists\ a_i,\ a_j \in \mathcal{B}$ *s.t.* $(a_i, a_j) \in \mathcal{R}$.
– $\mathcal{B}$ defends a_i *iff* $\forall\ a_j \in \mathcal{A}$, *if* $(a_j, a_i) \in \mathcal{R}$, *then* $\exists\ a_k \in \mathcal{B}$ *s.t.* $(a_k, a_j) \in \mathcal{R}$.

The main semantics introduced by Dung are summarized in the following definition.

Definition 3 (Acceptability semantics [6]) *Let $\mathcal{B}$ be a conflict-free set of arguments.*

– $\mathcal{B}$ *is* admissible *iff it defends any argument in $\mathcal{B}$.*
– $\mathcal{B}$ *is a* preferred *extension iff it is a maximal (w.r.t $\subseteq$) admissible extension.*
– $\mathcal{B}$ *is a* stable *extension iff it is a preferred extension that attacks any argument in $\mathcal{A}\backslash\mathcal{B}$.*

Example 1 *Let $T = (\mathcal{A}, \mathcal{R})$ be an argumentation theory where $\mathcal{A} = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ is the set of the arguments and $\mathcal{R} = \{(a_1, a_2), (a_2, a_1), (a_1, a_4), (a_2, a_3)\}$ is the set of attacks. This argumentation theory has two stable extensions $\mathcal{E}_1 = \{\alpha_1, \alpha_3\}$ and $\mathcal{E}_2 = \{\alpha_2, \alpha_4\}$*

3 Preference-based Argumentation Framework: Properties and Limitations

In [10] the basic argumentation framework of Dung has been extended into *preference-based argumentation theory (PBAT)*. The framework has been further developed and studied in [7]. The basic idea of a PBAT is to consider two binary relations between arguments:

1. A *conflict* relation, denoted by $\mathcal{C}$, that is based on the logical links between arguments.
2. A *preference* relation, denoted by $\succeq$, that captures the idea that some arguments are stronger than others. Indeed, for two arguments $a, b \in \mathcal{A}$, $a \succeq b$ means that a is at least as good as b. The relation $\succeq$ is assumed to be a partial pre-order (that is *reflexive* and *transitive*). The relation $\succ$ denotes the corresponding strict relation. That is, $a \succ b$ iff $a \succeq b$ and $b \not\succeq a$.

The two relations are combined into a unique attack relation, denoted by $\mathcal{R}$, and the Dung's semantics are applied on the resulting framework. In what follows, we focus on a particular class of PBATs, presented in [7], where the conflict relation $\mathcal{C}$ is *irreflexive* and *symmetric*.

Definition 4 (Preference-based Argumentation Theory (PBAT)) *([3]) Given an irreflexive and symmetric conflict relation $\mathcal{C}$ and a preference relation $\succeq$ on a set of arguments $\mathcal{A}$, a* preference-based argumentation theory *(*PBAT*) on $\mathcal{A}$ is an argumentation system $T = (\mathcal{A}, \mathcal{R})$, where $(a, b) \in \mathcal{R}$ iff $(a, b) \in \mathcal{C}$ and $b \not\succ a$.*

It follows directly from the definition that if $(a, b) \in \mathcal{C}$ and $a \succeq b$ and $b \not\succeq a$, then $(a, b) \in \mathcal{R}$. Moreover, if $(a, b) \in \mathcal{C}$ and a, b are either indifferent or incompatible in $\succeq$, then $(a, b) \in \mathcal{R}$ and $(b, a) \in \mathcal{R}$. Also note that if $(a, b) \in \mathcal{C}$, then either $(a, b) \in \mathcal{R}$ or $(b, a) \in \mathcal{R}$. Finally, if $(a, b) \in \mathcal{R}$ and $(b, a) \notin \mathcal{R}$, then $a \succ b$.

The following example illustrates some features of PBATs.

Example 2 *Let $\mathcal{A} = \{a, b, c, d\}$ be a set of arguments, and $\mathcal{C}$ the conflict relation on $\mathcal{A}$ defined as $\mathcal{C} = \{(a, b), (b, a), (b, c), (c, b), (c, d), (d, c)\}$. Moreover, let the preference relation $\succeq$ contain transitive closure of the set of pairs $a \succeq b$, $b \succeq c$, $c \succeq d$, and $d \succeq c$. The corresponding PBAT is $T = (\mathcal{A}, \mathcal{R})$, where $\mathcal{R} = \{(a, b), (b, c), (c, d), (d, c)\}$. Theory T has two stable extensions, $E_1 = \{a, c\}$ and $E_2 = \{a, d\}$.*

In [3] the impact of the preference relation on an argumentation system has been studied. After defining a relation $\rhd$ on the powerset $2^{\mathcal{A}}$ of the arguments of a PBAT $T = (\mathcal{A}, \mathcal{R})$, it has been shown that the stable extensions of T correspond to the most preferred elements of $2^{\mathcal{A}}$ wrt this relation.

Definition 5 *([3]) Let $T = (\mathcal{A}, \mathcal{R})$ be a PBAT built on an underlying pre-order $\succeq$. If $A_1, A_2 \in 2^{\mathcal{A}}$, with $A_1 \neq A_2$, then $A_1 \rhd A_2$ iff one of following holds:*

- *$A_1 \supset A_2$*
- *for all a, b such that $a \in A_1 \setminus A_2$ and $b \in A_2 \setminus A_1$, it holds that $a \succ b$*

The exact correspondence between the relation $\rhd$ and stable extensions is as follows.

Theorem 1 *([3]) Let $T = (\mathcal{A}, \mathcal{R})$ be a PBAT built on an underlying pre-order $\succeq$ and a conflict relation $\mathcal{C}$. E is a stable extension of T iff there are no arguments $a, b \in E$ s.t. $(a, b) \in \mathcal{C}$, and for all $A \in 2^{\mathcal{A}}$ such that $A \rhd E$, there are $a_1, a_2 \in A$ such that $(a_1, a_2) \in \mathcal{C}$.*

The example below explains the link between $\rhd$ and stable extensions.

Example 3 *Let $T = (\mathcal{A}, \mathcal{R})$ be a PBAT with $\mathcal{A} = \{a, b, c\}$ and $\mathcal{R}$ composed from the conflict relation $\mathcal{C} = \{(a, b), (b, a)(a, c), (c, a)\}$ and preference relation that contains the pairs $a \succ b$ and $a \succ c$, and marks all other pairs of arguments as indifferent. The relation $\rhd$ on $2^{\mathcal{A}}$ induced by $\succeq$ contains the pairs $\{a\} \rhd \{b, c\}$, $\{a\} \rhd \{b\}$, $\{a\} \rhd \{c\}$. Since the sets $\{a, b, c\}$, $\{a, b\}$, $\{a, c\}$ are ruled out by $\mathcal{C}$, the set $E = \{a\}$ is the stable extension of T.*

One feature of the $\rhd$ relation is that it may *not be transitive*. Consider for instance the theory of the previous example, and observe that $\{a, b\} \rhd \{b, c\} \rhd \{c\}$. However, it is not the case that $\{a, b\} \rhd \{c\}$.

The second important observation, which is the main focus of this work, relates to the conclusions sanctioned by preference-based argumentation under the stable model semantics. The following example, borrowed from [8], shows clearly that these results can be counterintuitive even in simple cases.

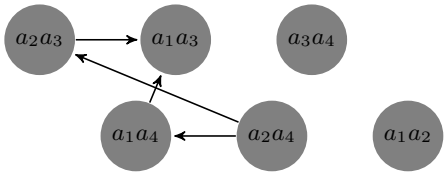

Fig. 1. The preference relation induced on sets of arguments.

Example 4 *The story of the example is about conclusions that can been drawn regarding the financial situation of a person based on arguments built on information about her occupation and residence. Let's suppose that layers are, in general, considered to be wealthy, but a certain subclass, the pubic defenders, are considered not to be. Consider now an area in Paris -say, Passy - containing a large number of expensive private homes along with a much smaller number of middle-income rental properties. Thus the residents of Passy can be generally considered to be wealthy although the renters, are considered to not to be. Assume that Ann is a public defender (PDa), and therefore a layer (La), who rents in Passy (Ra), and is therefore a resident of Passy (Pa).*

If we assume that Wa represents the proposition that Ann is wealthy, the arguments that can be generated in an underlying propositional language from the above story are $a_1 = \{PDa, PDa \longrightarrow La, La \longrightarrow Wa\}$, $a_2 = \{PDa, PDa \longrightarrow \neg Wa\}$ $a_3 = \{Ra, Ra \longrightarrow Pa, Pa \longrightarrow Wa\}$, $a_4 = \{Ra, Ra \longrightarrow \neg Wa\}$.

From the above arguments we generate the PBAT $T=(\mathcal{A}, \mathcal{R})$, *where* $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$. *The attack relation* $\mathcal{R}$ *is composed from the conflict relation* $C = \{(a_1, a_2), (a_2, a_1), (a_1, a_4), (a_4, a_1), (a_3, a_2), (a_2, a_3), (a_3, a_4), (a_4, a_3)\}$, *and the preference relation* $\succeq$ *that is defined as* $a_2 \succ a_1$ *and* $a_4 \succ a_3$, *whereas all other pairs of arguments are incomparable.*

Theory T has two extensions, namely $E_1 = \{a_1, a_3\}$ *and* $E_2 = \{a_2, a_4\}$. *The first extension supports the conclusion that Ann is wealthy, whereas the the second that she is not. Intuitively however one would conclude that Ann is not wealthy. In other words we could argue that the second extension is more preferred than the first, as for every argument of the first it contains a more preferred argument.*

The relation $\rhd$ *on the subset of* $\mathcal{A}$ *with two elements is depicted in figure 3. Note again that* $\rhd$ *is not transitive. Indeed, it holds that* $\{a_2, a_4\} \rhd \{a_2, a_3\}$ *and* $\{a_2, a_3\} \rhd \{a_1, a_3\}$, *but* $\{a_2, a_4\} \not\rhd \{a_1, a_3\}$. *Although,* $a_2 \succ a_1$ *and* $a_4 \succ a_3$, *the stable model semantics does not render* $\{a_2, a_4\}$ *better than* $\{a_1, a_3\}$, *because* $a_2 \not\succ a_3$ *and* $a_4 \not\succ a_1$.

The main purpose of this work is to provide a preliminary study of the problem of the conclusions sanctioned by the state-of-the-art argumentation, and identify possible solutions by borrowing ideas from decision theory.

4 Preference-based Argumentation revisited

As we noted in section 3, the stable model semantics can lead to counter-intuitive results. To remedy the situation we present a new semantics for preference-based argumentation called *super-stable* extensions semantics. The main idea is to only accept conclusions drawn under the stable model semantics from a PBAT T that correspond to the conclusions that are sanctioned by some other PBAT T' which is obtained from T by removing incomparability. Therefore, the new semantics may differ from the standard stable model semantics only on theories with incomparability. As we will discuss in the next section theories without incomparability always sanction the correct conclusion under the stable extension semantics.

Before we proceed to the definition of the new semantics, we recall some useful concepts. A relation $\succeq$ on a set S is *total* if for all $a, b \in S$ with $a \neq b$, $a \succeq b$ or $b \succeq a$. A *strict total order* on a set S is an asymmetric (hence irreflexive), transitive and total relation on S. The notion of an *extension* of a relation is used in decision theory and economics eg. ([11], [12])

Definition 6 *A binary relation $\succeq_E$ on S is an extension of a pre-order $\succeq$ on S if $\succeq_E$ is a pre-order on S such that $\succeq_E \supseteq \succeq$ and for all $a, b \in S$ if $a \succ b$ then $a \succ_E b$. An extension of a pre-order $\succeq$ that is complete (ie., for all $a, b \in S$, $a \succeq b$ or $b \succeq a$) is called ordering extension of $\succeq$.*

Hansson [11] has shown that every pre-order has an ordering extension. Moreover, Donaldson and Weymark [12] proved that a pre-order is the intersection of its ordering extensions.

Definition 7 *A strict total order $\succ_s$ on a set S is a strict ordering of a total pre-order $\succeq$ if for all $a, b \in S$ if $a \succ b$ then $a \succ_s b$.*
A strict total order is a strict ordering of a pre-order if it is a strict ordering of one of its ordering extensions.

The following definition extends the notions of ordering extension and strict ordering to the case of PBATs.

Definition 8 *Let $T = (\mathcal{A}, \mathcal{R})$ be a PBAT on an underlying pre-order $\succeq$ and a conflict relation $\mathcal{C}$. The PBAT $T_o = (\mathcal{A}, \mathcal{R}_o)$, on an underlying relation $\succeq_o$ and the conflict relation $\mathcal{C}$, is a* ordering completion *of T if $\succeq_o$ is an ordering extension of $\succeq$. The PBAT $T_s = (\mathcal{A}, \mathcal{R}_s)$, on an underlying relation $\succ_s$ and the conflict relation $\mathcal{C}$, is a* strict projection *of T if $\succ_s$ is a strict ordering of $\succeq$.*

The following result that relates the stable extensions of a PBAT and the stable extensions of its ordering completions and strict projections is easily provable.

Proposition 1 *Let T be a PBAT, T_o one of its ordering completions, and T_s one of its strict projections. If E_o is a stable extension of T_o, then it is also a stable extension of T. Moreover, if E_s is a stable extension of T_s, then it is also a stable extension of T.*

However, it is not the case that a stable extension of T is a stable extension of some T_o built on a pre-order $\succ_o$ that is an ordering extension of $\succeq$. Consider for instance again the theory of example 4, and its stable extension $E_1 = \{a_1, a_3\}$. To see that there is no PBAT T_o which is an ordering completion of T and has E_1 as a stable extension, observe that for E_1 to be a stable extension it must be the case that $a_1 \succeq a_4$ and $a_3 \succeq a_2$. However, together with $a_2 \succ a_1$ these would mean that $a_3 \succeq a_4$, which is impossible given that $a_4 \succ a_3$.

We now proceed with the definition of the new semantics for preference-based argumentation. As noted earlier, the basic idea is to only accept a set of arguments as an extension of a PBAT T if this set is a stable extension of an ordering completion of T. More formally the concept is defined as follows.

Definition 9 *Let $T = (\mathcal{A}, \mathcal{R})$ be a PBAT built on an underlying pre-order $\succeq$ and conflict relation $\mathcal{C}$. A stable extension E of T is a* super-stable *extension of T if it is the stable extension of an ordering extension of T.*

By the results of [11] we know that every pre-order has an ordering extension. Therefore, every PBAT has an ordering completion which is itself a PBAT. From [7], we know that every PBAT has a stable extension. By combining these two results we obtain the following property for super-stable extensions.

Proposition 2 *Every PBAT has a super-stable extension.*

5 Theories without incomparability

If a PBAT T contains no incomparability, T is an ordering completion of itself. Therefore, any stable extension of T is by definition a super-stable extension of T. In this section we prove that for this class of theories a correspondence holds between their stable extensions and the stable extensions of their strict projections.

For a PBAT without incomparability, $T = (\mathcal{A}, \mathcal{R})$, we define the *level* of argument $a \in \mathcal{A}$, denoted by $l(a)$, recursively as follows

- $l(a) = 1$ for all a such that there is no $b \in \mathcal{A}$ s.t. $b \succ a$
- $l(a) = k$ for all a such that for all $a' \in \mathcal{A}$ s.t. $a' \succ a$ it holds that $l(a') < k$, and $\exists a'' \in \mathcal{A}$ s.t. $a' \succ a$ and $l(a') = k - 1$

The following lemma relates the level of arguments with relation $\succ$ and will be used in the proof of the main result of this section (proposition 3 below).

Lemma 1. *Let $T = (\mathcal{A}, \mathcal{R})$ be a PBAT without incomparability on an underlying pre-order $\succeq$. For every $a, b \in \mathcal{A}$, $l(a) < l(b)$ iff $a \succ b$.*

Proof. The property that for every $a, b \in \mathcal{A}$ if $a \succ b$ then $l(a) < l(b)$, follows directly from the definition of the level of an argument. We prove now that if $l(a) < l(b)$ then $a \succ b$. Assume $a, b \in \mathcal{A}$ with $l(a) < l(b)$. Clearly it can not be the case $b \succ a$, because then $l(b) < l(a)$. Assume that $a \succeq b$ and $b \succeq a$, and

$l(b) = m$ (hence, $l(a) < m$). Then, there must be $c \in \mathcal{A}$ s.t $l(c) = m - 1$ and $c \succ b$. Therefore, it must be the case that $c \succ a$, and hence $l(a) \geq m$. But this contradicts $l(a) < m$. Therefore it holds that $a \succ b$. □

A direct consequence of the previous lemma is that $l(a) \leq l(b)$ iff $a \succeq b$. Also note that since a super-stable extension of a PBAT $T = (\mathcal{A}, \mathcal{R})$ is also a stable extension of T, it holds that for all $a \notin E$ there exist $b \in E$ such that $(b, a) \in \mathcal{R}$. From the above we conclude that for all $a \notin E$ there exist $b \in E$ such that $(a, b) \in \mathcal{C}$ and $l(b) \leq l(a)$. Note the above property does not hold in general for theories that contain incomparability.

Proposition 3 *Let T be a PBAT without incomparability. Every stable extension of T is a stable extension of some strict projection of T.*

Proof. We prove the claim by defining a strict projection of T, $T_s = (\mathcal{A}, \mathcal{R}_s)$, for which E is a stable extension. Let $\succ_s^{E^+}$ be a strict projection on the arguments of E. Similarly, let $\succ_s^{E^-}$ be a strict projection on the arguments of $\mathcal{A} \setminus E$. Finally, let $\succ_s^{E^+,-}$ be the binary relation on $(E \times (\mathcal{A} \setminus E)) \cup ((\mathcal{A} \setminus E) \times E)$ such that for any pair of arguments $a \in E$ and $b \notin E$

- if $l(a) > l(b)$ then $b \succ_s^{E^+,-} a$
- if $l(a) \leq l(b)$ then $a \succ_s^{E^+,-} b$

where $l(a)$ is the level of argument a in theory T. Define $\succ_s = \succ_s^{E^+} \cup \succ_s^{E^-} \cup \succ_s^{E^+,-}$.
We first show that $\succ_s$ is strict total order. It is easy to verify that $\succ_s$ is asymmetric and total by construction. We show that $\succ_s$ is transitive. Let $a, b, c \in \mathcal{A}$, such that $a \succ_s b$ and $b \succ_s c$. We need to show that $a \succ_s c$. We proceed by case analysis. For the case where $a, b, c \in E$ or $a, b, c \in \mathcal{A} \setminus E$, transitivity follows by construction.
Assume now that $a \in E$ and $b, c \notin E$. Then, by construction, $l(a) \leq l(b)$. Moreover, since $b \succ_s c$, by lemma 1, it must be the case $l(b) < l(c)$. Hence, $l(a) < l(c)$, which, again by lemma 1, means that $a \succ_s c$. The case where $a \in E$, $b \notin E$, $c \in E$ is similar.
Now suppose that $a, b \in E$, and $c \notin E$. Since $a \succ_s b$, by lemma 1, we obtain that $l(a) < l(b)$. Moreover, by construction, it must hold that $l(b) \leq l(c)$. Hence, $l(a) < l(c)$, and therefore $a \succ_s c$.
The remaining cases where $a \notin E$ can be proved analogously.

Finally, we prove that E is a stable extension of $T_s = (\mathcal{A}, \mathcal{R}_s)$. First note that E is conflict free. Now assume that $a \notin E$ and define $D(a) = \{b | b \in E \text{ and } (b, a) \in \mathcal{R}\}$. Since E is a stable extension, it must be $D(a) \neq \emptyset$. Let $b \in D(a)$, and assume that $(b, a) \notin \mathcal{R}_s$. By construction, it must hold that $l(a) < l(b)$, which by lemma 1 implies $a \succ b$. This however contradicts $b \in D(a)$. Therefore, $(b, a) \in \mathcal{R}_s$, which means that E is a stable extension of T_s. □

By combining the definition of a super-stable extension with proposition 3 we obtain the following strong property regarding super-stable extensions.

Proposition 4 *Every super-stable extension of a PBAT T is a stable extension of some strict projection of T.*

6 A multi-criteria view of PBAT

In the previous sections we investigated how standard argumentation semantics can be extended to accommodate preference information on the arguments. In this section we change our perspective and explore a more direct link between argumentation and decision theory. More specifically, we interpret arguments as criteria and regard preferences as information on the relative importance of these criteria. Under this perspective argumentation can be understood as Multiple Criteria Decision Problem (MCDP). We start our analysis with a definition of the problem that leaves out some of its aspects that are not directly relevant to our purposes.

Definition 10 *A* Multiple Criteria Decision Problem (MCDP) *is a triple $P = (A, K, \succeq)$ where*

- $A = \{a_1, \ldots, a_n\}$ *is the set of* attributes.
  A set of values is associated with each attribute, denoted by $v(a_1), \ldots, v(a_n)$
- $K = \{\gg_{a_1}, \ldots, \gg_{a_n}\}$ *is the set of* criteria. *A criterion $\gg_{a_i}$ is a pre-order associated with the values of an attribute a_i*
- $\succeq$ *is pre-order on the criteria*

An alternative *l wrt to a MCDP $P = (A, K, \succeq)$ is any $l \in v(a_1) \times \ldots \times v(a_n)$. We denote the set of alternatives by L_P.*

In certain situations, a *solution* to a MCDP is a ranking relation $\trianglerighteq$ on the set of alternatives L, ie. $\trianglerighteq \subset L \times L$. Usually a solution to a MCDP has to satisfy certain properties [13].

The following definition shows that argumentation can be transformed in a meaningful way into a MCDP.

Definition 11 *Given a PBAT $T = (\mathcal{A}, \mathcal{R})$, where $\mathcal{A} = \{a_1, \ldots, a_n\}$, we define its corresponding MCDP $M_T = (A_T, K_T, \succeq_T)$ as follows:*

- $A_T = \mathcal{A}$, *with $v(a_i) = \{a_i^+, a_i^-\}$, for each $a_i \in A_T$.*
- $K_T = \{\gg_1, \ldots, \gg_n\}$, *where $\gg_i$, for $1 \leq i \leq n$, is defined as the preference $a_i^+ \gg_i a_i^-$.*
- $\succeq_T = \succeq$

The following is an example of a translation of a specific PBAT into a MCDP.

Example 5 *Consider the PBAT $T = (\mathcal{A}, \mathcal{R})$, where $\mathcal{A} = \{a_1, a_2, a_3\}$, and the underlying preference relation $\succeq$ defined as: $a_1 \succ a_2$, $a_1 \succ a_3$, $a_2 \succeq a_3$, $a_3 \succeq a_2$. The corresponding MCDP is defined as $M_T = (A_T, K_T, \succeq_T)$, where:*

– $A_T = \{a_1, a_2, a_3\}$, *with* $v(a_1) = \{a_1^+, a_1^-\}$, $v(a_2) = \{a_2^+, a_2^-\}$, $v(a_3) = \{a_3^+, a_3^-\}$.
– $K_T = \{\gg_1, \gg_2, \gg_3\}$, *with* $a_1^+ \gg_1 a_1^-$, $a_2^+ \gg_2 a_2^-$, $a_3^+ \gg_3 a_3^-$.
– $\succeq_T = \succeq$

Several methods have been proposed in the literature, the applicability of which in many cases depends on the features of the MCDP at hand. Among the methods for tackling MCDPs that appear in the literature and are applicable to the case we consider, it seems that the Regime method [9] is the closest to the spirit of the stable extensions semantics. Here we discuss a simplified version of the method as it appears in [14].

The Regime method works as follows. For any two alternatives A_i, A_j, let K^+ be the set of criteria according to which A_i is better than A_j, and K^- be the set of criteria according to which A_j is better than A_i. Regime ranks A_i better than A_j, denoted by $A_i \rhd_R^t A_j$, if $K^+ \neq \emptyset$ and there is an injective map from K^- to K^+ by which each criterion in K^- is mapped to a more important criterion in K^+. The set of optimal alternatives is then $\{A_o : \forall i \neg(A_i \rhd_R^t A_o)\}$.

We can easily define a preference order on the sets of arguments of a PBAT that captures the Regime method. To do this we associate to any set of arguments A, an alternative $A^t = \{a^+ | a \in A\} \cup \{a^- | a \notin A\}$.

Definition 12 *Let $T = (\mathcal{A}, \mathcal{R})$ be a PBAT and $M_T = (A_T, K_T, \succeq_T)$ its corresponding MCDP. For any $A_1, A_2 \subseteq 2^{\mathcal{A}}$, it holds that $A_1 \rhd_R A_2$ if $A_1^t \rhd_R^t A_2^t$.*

We can now define the notion of a *regime extension* of a PBAT by characterizing it in a way similar to the stable extensions.

Definition 13 *A set of arguments E is a* regime extension *of a PBAT $T = (\mathcal{A}, \mathcal{R})$ if there are no arguments $a, b \in E$ s.t. $(a, b) \in \mathcal{C}$, and for all $E' \in 2^{\mathcal{A}}$ such that $E' \rhd_R E$, there are $a_1, a_2 \in E'$ such that $(a_1, a_2) \in \mathcal{C}$.*

We now apply the previous definition to the story of 4, and observe that it yields the correct result.

Example 6 *Consider again the theory T of example 4. The corresponding MCDP $M_T = (A_T, K_T, \succeq_T)$ can be defined as outlined above. Consider the two sets of arguments $E_1 = \{a_1, a_3\}$ and $E_2 = \{a_2, a_4\}$ of T which correspond to the alternatives $E_1^t = \{a_1^+, a_2^-, a_3^+, a_4^-\}$ and $E_2^t = \{a_1^-, a_2^+, a_3^-, a_4^+\}$. For the comparison $E_2^t \rhd_R^t E_1^t$ we have that $K^+ = \{a_2, a_4\}$, $K^- = \{a_1, a_3\}$, and the mapping $a_1 \rightarrow a_2$, $a_3 \rightarrow a_4$. For the comparison $E_1^t \rhd_R^t E_2^t$ we have that $K^+ = \{a_1, a_3\}$ $K^- = \{a_2, a_4\}$, but there is no suitable mapping. Therefore we conclude $E_2 \rhd_R E_1$. Moreover, E_2 is a regime extension of T.*

7 Conclusion and Future Work

In this paper we pointed out that Dung's stable extensions semantics when applied in preference-based argumentation frameworks for decision making purposes lead to counter-intuitive conclusions. A similar problem has been identified

also by Horty in [8]. This mainly holds for argumentation theories where the preference relation used for defining the relative strength of individual arguments contains incomparability. To resolve this problem we proposed a new acceptability semantics called *super-stable extensions* which allows to capture the conclusions corresponding to the good decisions and to avoid the counter intuitive ones which could correspond to bad decisions. Moreover, we showed that preference-based argumentation can be understood as a multiple-criteria decision problem allowing to that way the exploration of the application of theoretical results of the decision theory in argumentation. Therefore, this work can been seen as an attempt to bring new ideas from decision theory to argumentation.

Our future work concerns the definition of a binary relation $\rhd_{SS}$ on the sets of arguments of a PBAT that will be proved exactly the preference relation that is induced by the super-stable extensions semantics and to prove the correspondence between both. This will be the equivalent result of the one we proved in [7] between the preference $\rhd$ and the stable extensions.

References

1. Kakas, A., Moraitis, P.: Argumentation based decision making for autonomous agents. In: Proc. 2nd International Joint Conference on Autonomous Agents and Multi-Agents systems. (2003) 883–890
2. Amgoud, L.: A general argumentation framework for inference and decision making. In: 21st Conference on Uncertainty in Artificial Intelligence, UAI'2005. (2005) 26–33
3. Amgoud, L., Dimopoulos, Y., Moraitis, P.: Making decisions through preference-based argumentation. In: Principles of Knowledge Representation and Reasoning, KR. (2008) 113–123
4. Bonet, B., Geffner, H.: Arguing for decisions: A qualitative model of decision making. In: Proceedings of the 12th Conference on Uncertainty in Artificial Intelligence. (1996) 98–105
5. Amgoud, L., Prade, H.: Explaining qualitative decision under uncertainty by argumentation. In: 21st National Conference on Artificial Intelligence, AAAI'06. (2006) 16 – 20
6. Dung, P.M.: On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. Artificial Intelligence **77** (1995) 321–357
7. Dimopoulos, Y., Moraitis, P., Amgoud, L.: Theoretical and Computational Properties of Preference-based Argumentation. In: Proc. of European Conference on AI (ECAI'08). (2008)
8. Horty, J.: Argument construction and reinstatement in logics for defeasible reasoning. Artificial Intelligence and Law **9** (2001) 1–28
9. Hinloopen, E., Nijkamp, P., Rietveld, P.: The regime method: A new multicriteria technique. In Hansen, P., ed.: Essays and Surveys on Multiple Criteria Decision Making. Springer (1983) 146–155
10. Amgoud, L., Cayrol, C.: On the acceptability of arguments in preference-based argumentation framework. In: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence. (1998) 1–7

11. Hansson, B.: Choice structures and preference relations. Synthése **18** (1968) 443–458
12. Donaldson, D., Weymark, A.: A quasiordering is the intersection of orderings. Journal of Economic Theory **178** (1998) 382–387
13. Bouyssou, D., Marchant, T., Pirlot, M., Perny, P., Tsoukias, A., Vincke, P.: Multicriteria Evaluation and Decision Models: stepping stones for the analyst. Springer Verlag (2006)
14. Moffett, A., Sarkar, S.: Incorporating multiple criteria into the design of conservation area networks: A minireview with recommendations. Diversity and Distributions **12** (2006) 125–137