Characterizing the Outcomes of Argumentation-based Integrative Negotiation

Yannis Dimopoulos[1] Pavlos Moraitis[2]
Leila Amgoud[3]
[1] University of Cyprus, 75 Kallipoleos Str. 1678, Nicosia-Cyprus
[2] Paris Descartes University, 45 rue des Saints-Pères 75270, Paris-France
[3] Paul Sabatier University, 118 route de Narbonne 31062, Toulouse-France

Abstract

In the negotiation literature we find two relatively distinct types of negotiation. The two types are known as integrative negotiations *and* distributive negotiations*. Integrative negotiations are those where all sides are looking for solutions that are "good" for everyone while distributive negotiations are those where each party tries to maximize his gain. In this paper we are interested in argumentation-based integrative negotiations. More precisely we present a study characterizing the outcomes of such negotiations. For this reason, we aggregate the argumentation systems that the agents use in order to negotiate. The aggregate argumentation system represents the negotiation theory of the agents as a group and corresponds to the "ideal" situation of having access to complete information or negotiating through a mediator. We show that the aggregation operator we use is very suitable for capturing the essence of integrative negotiation as the outcomes of the aggregate theory we obtain have many appealing properties (e.g. they are Pareto optimal solutions).*

1 Introduction

Negotiation is an important issue in multi-agent systems (MAS) field. Different approaches to automated negotiation have been investigated in MAS [7], including *game-theoretic* approaches [10], *heuristic-based* approaches [7], and *argumentation-based* approaches (see e.g. [9])

In the "classical" negotiation literature we find two relatively distinct types of negotiation. The two types are known as *integrative negotiations* and *distributive negotiations* [6]. Integrative negotiations are those where all sides are looking for a solution that allows everyone to walk away feeling like they won something. Distributive negotiations, on the other hand, are typically those where one party gets what he wants while the other party gives something up.

In this paper we are interested in *argumentation-based integrative negotiations*. More precisely, we consider negotiations that takes place between two (or more) agents on a set $\mathcal{O}$ of offers (or outcomes), whose structure is not known. The offers may concern one or several issues. However the way that issues are fixed or combined during the negotiation is a matter of strategy and it is out of the scope of this paper. The goal of the negotiation is to find among elements of $\mathcal{O}$, an offer that satisfies more or less the preferences of both agents. Each agent is supposed to have a *negotiation theory* represented in an abstract way. A negotiation theory consists of a set $\mathcal{A}$ of arguments whose structure and origin are not known, a function specifying for each possible offer (or outcome) in $\mathcal{O}$, the arguments of $\mathcal{A}$ that support it, a non specified conflict relation among the arguments, and finally a preference relation between the arguments. The status of each argument is defined using Dung's acceptability semantics [5].

Our aim is to characterize the outcomes of such negotiations. In order to have the complete picture of the different possible outcomes we may obtain in such a framework, we study the "ideal" situation which assumes complete information between the negotiating agents. This could be also correspond to the situation where a *mediator* receives the arguments of all the involved agents and tries to find the best compromise for them. For this reason, we propose a characterization of the outcomes which is based on the aggregation of the individual negotiation theories and which generates a single theory that contains the arguments that the agents have as a group. This allows a reasoning mechanism to identify the outcomes that are "good" for all the negotiating agents.

Thus, more concretely, we assume that each agent has a negotiation theory and an argumentation-based reasoning mechanism as they are originally proposed in [2] and further extended in [3]. Then we propose an aggregation of the negotiation theories in order to study the possible outcomes. The *aggregation operator* we are using combines in a simple and intuitive way the preferences of the different agents. More specifically, the preference relation of the aggregate theory compares only those pairs of arguments for which the agents' preferences agree (i.e. they have the same preference or they are indifferent), and regards as incomparable the ones on which they disagree (i.e. they have opposite preferences).

It turns out that the outcomes of the aggregate theory have interesting properties. More precisely, under the preferred extensions semantics, the retained arguments (and therefore the offers they support) are not "dominated" by any other argument wrt the aggregate preference relation. These arguments are exactly the *Pareto optimal* arguments. That means that the outcomes (i.e. the agreed offers) of the negotiation are Pareto optimal solutions for the agents. More interestingly, there are arguments that belong to the stable extensions of the aggregate theory but do not belong to any of the stable extensions of the individual theories. Therefore, that means that it is beneficial for the agents to negotiate, since this may reveal further outcomes that are of mutual interest.

The rest of the paper is organized as follows. The next section introduces the argumentation system used by the agents for negotiating. Then, we present the aggregate system and its formal properties. Finally, we conclude with some remarks and perspectives. For space reasons the proofs of the presented theoretical results can be found in www.math-info.univ-paris5.fr/∼moraitis/webpapers/IAT08-proofs

2 The Argumentation System

The negotiation theories of the agents are represented as preference-based argumentation theories and are used in the negotiation process by an argumentation system introduced in [2] and further developed in [3]. It is based on two disjoint sets of arguments, one for supporting *beliefs* and one that supports *offers* or *options*. Here we concentrate on the analysis of argument structures supporting offers, and ignore completely arguments that refer to beliefs. Nevertheless, it should be noted that the results we present here carry over to the general case, with slight modifications.

In order this work to be self-contented we present here the basic elements of the negotiation theories. In what follows, $\mathcal{L}$ will denote a logical language, and $\equiv$ is an equivalence relation associated with it. From $\mathcal{L}$, a set $\mathcal{O}$ = $\{o_1, \ldots, o_n\}$ of n *offers* is identified, such that $\nexists o_i, o_j \in \mathcal{O}$ such that $o_i \equiv o_j$. This means that the offers are distinct. Different *arguments* can be built from $\mathcal{L}$ for supporting elements of $\mathcal{O}$. The set $\mathcal{A}$ will contain all those arguments. By argument, we mean a *reason* for supporting an offer. The structure and origin of the arguments are assumed to be unknown. Let $\mathcal{F}$ be a function that returns for a given offer $o \in \mathcal{O}$, the arguments in favor of it. Thus, $\mathcal{A}$ = $\bigcup_{i=1,\ldots,n} \mathcal{F}(o_i)$. There are two relations among arguments:

1. A conflict relation, denoted by $\mathcal{R}$, that is based on the logical links between arguments. This relation is assumed to be *weakly complete*, that is *irreflexive*, *symmetric* and *complete* in the sense that any two distinct arguments are related by $\mathcal{R}$. In fact, for two arguments $a, b \in \mathcal{A}$, $(a, b) \in \mathcal{R}$ means that a and b are conflicting.

The arguments of the set $\mathcal{A}$ are therefore pairwise conflicting, representing situations where each agent has to decide in favor of *one* outcome i.e. the best for him and abandon the others. These situations are abundant in integrative negotiation where agents must at each time be able to decide about what they must defend and what they must abandon in order to reach a compromise. Therefore, the arguments that support the different possible alternative offers they may do are essentially pairwise conflicting. Arguments supporting the same offer are also considered as conflicting. The reason behind this idea is the requirement that the system should return at each time the most powerful argument among the several ones that support the offer under consideration.

2. A preference relation, denoted by $\succeq$, capturing the idea that some arguments are stronger than others. Indeed, for two arguments $a, b \in \mathcal{A}$, $(a, b) \in \succeq$ means that a is at least as good as b. At some places, we will also write $a \succeq b$. In what follows, the relation $\succeq$ is assumed to be a partial pre-ordering (that is *reflexive* and *transitive*). The relation $\succ$ denotes the strict partial order associated with $\succeq$. It is defined as follows: $(a, b) \in \succ$ iff $(a, b) \in \succeq$ and $(b, a) \notin \succeq$.

The formal definition of a weakly complete relation is a follows.

Definition 1 (Weakly complete relation). *Let $\mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. The relation $\mathcal{R}$ is* weakly complete *iff $\mathcal{R}$ = $\{(x, y) | \; x, y \in \mathcal{A}$ and $x \neq y\}$.*

Before we proceed, let us start by stating some useful properties of the preference relation $\succeq$, that will be used throughout the paper.

Property 1. *Let $a, b \in \mathcal{A}$ such that $a \succeq b$ and $b \succeq a$. Then, the following hold.*

1. *If $a \succeq c$, for some $c \in \mathcal{A}$, then $b \succeq c$.*
2. *If $c \succeq a$, for some $c \in \mathcal{A}$, then $c \succeq b$.*
3. *If $a \not\succeq c$ and $c \not\succeq a$ for some $c \in \mathcal{A}$, then $b \not\succeq c$ and $c \not\succeq b$.*

The preference relation and the conflict relation are combined into a unique relation between arguments, denoted by `Def` and called *defeat* relation, as in [1].

Definition 2 (Defeat relation). *Let $\mathcal{A}$ be a set of arguments, $\succeq$ a pre-order on $\mathcal{A}$, and $a, b \in \mathcal{A}$. a `Def` b iff:*

- *$a \; \mathcal{R} \; b$, and*

- *not* ($b \succeq a$)

Note that when the preference relation $\succeq$ is empty, i.e. all the arguments are incomparable, then the relation `Def` coincides with $\mathcal{R}$. Moreover, in the case of weakly complete systems, the relation `Def` is exclusively induced by the preference relation $\succeq$. Namely, `Def` can be defined in terms of $\succeq$ as $\texttt{Def} = ((\succeq)^C)^{-1}$, where S^C is the complement of a binary relation S, and S^{-1} its inverse, defined as $S^C = \{(x,y)|(x,y) \notin S\}$, and $S^{-1} = \{(x,y)|(y,x) \in S\}$. While, in principle, this seems to suggest that the conflict relation is not needed in the framework, it remains for two reasons. Firstly, it more clearly illustrates the link between the preference relation and the `Def` relation. Secondly, the conflict relation plays a central role in the case where beliefs are included in the language (the conflict relation is not complete in that case), and therefore is an important ingredient of the complete framework.

The following result states some useful properties of the relation `Def` that follow from the above definition. These properties highlight the impact of the preference relation on `Def`. The table below summarizes these properties.

$a \quad b$	$a \longrightarrow b$	$a \longleftrightarrow b$
$a \succeq b$ and $b \succeq a$	$a \succ b$	not($a \succeq b$) and not($b \succeq a$)

Property 2. *Let a, $b \in \mathcal{A}$.*

- $(a,b) \in \texttt{Def}$ *and* $(b,a) \in \texttt{Def}$ *iff* $a \not\succeq b$ *and* $b \not\succeq a$.
- $(a,b) \in \texttt{Def}$ *and* $(b,a) \notin \texttt{Def}$ *iff* $a \succ b$.
- $(a,b) \notin \texttt{Def}$ *and* $(b,a) \notin \texttt{Def}$ *iff* $a \succeq b$ *and* $b \succeq a$.

Notice that indifference obliterates the conflict relation. Indeed, if two arguments a, b are indifferent, then $(a,b) \notin$ `Def` and $(b,a) \notin$ `Def`. Hence, in a sense, the conflict relation applies only to arguments that are of different strength or incomparable.

An argumentation system for *negotiation* is defined as follows:

Definition 3 (Argumentation system). *An* argumentation system *is a tuple $T = \langle \mathcal{O}, \mathcal{A}, \succeq, \texttt{Def} \rangle$, where $\mathcal{O}$ is a set of offers, $\mathcal{A}$ is a set of arguments that support these offers, $\succeq$ a pre-order on $\mathcal{A}$, and* `Def` *is a defeat relation defined form the weakly complete relation on $\mathcal{A}$ and $\succeq$, as described in Definition 2.*

To simplify the notation, we may omit set $\mathcal{O}$ and the preference relation from the definition of an argumentation system, and denote it simply by $T = \langle \mathcal{A}, \texttt{Def} \rangle$. We use the two notations interchangeably. When the simplified notation $T = \langle \mathcal{A}, \texttt{Def} \rangle$ is used, we refer to $\succeq$ as the underlying preference relation of `Def`.

Note that to each argumentation system is associated a directed graph whose nodes are the different arguments of $\mathcal{A}$ and the arcs represent the defeats, i.e. elements of `Def`.

Among all the arguments, it is important to know which arguments to keep for making decisions. In [5], 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 4 (Conflict-free, Defence [5]). *Let $T = \langle \mathcal{A}, \texttt{Def} \rangle$ be a argumentation system, and $\mathcal{B} \subseteq \mathcal{A}$.*

- *$\mathcal{B}$ is* conflict-free *iff there $\nexists$ a_i, $a_j \in \mathcal{B}$ such that $(a_i, a_j) \in \texttt{Def}$.*
- *$\mathcal{B}$* defends *an argument a_i iff $\forall$ $a_j \in \mathcal{A}$, if $(a_j, a_i) \in \texttt{Def}$, then $\exists a_k \in \mathcal{B}$ such that $(a_k, a_j) \in \texttt{Def}$.*

Definition 5 (Acceptability semantics [5]). *Let $T = \langle \mathcal{A}, \texttt{Def} \rangle$ be a argumentation system, $\mathcal{B} \subseteq \mathcal{A}$ a conflict-free set of arguments, and $\mathcal{T}$: $2^{\mathcal{A}} \rightarrow 2^{\mathcal{A}}$ a function such that $\mathcal{T}(\mathcal{B})$ = $\{a \mid \mathcal{B}$ defends $a\}$.*

- *$\mathcal{B}$ is a* complete *extension of T iff $\mathcal{B}$ = $\mathcal{T}(\mathcal{B})$.*
- *$\mathcal{B}$ is an* admissible *extension of T iff it defends any element in $\mathcal{B}$.*
- *$\mathcal{B}$ is a* preferred *extension of T iff $\mathcal{B}$ is a maximal (w.r.t set $\subseteq$) admissible extension.*
- *$\mathcal{B}$ is a* stable *extension of T iff $\mathcal{B}$ is a preferred extension that defeats any argument in $\mathcal{A} \backslash \mathcal{B}$.*
- *$\mathcal{B}$ is a* grounded *extension of T iff it is the smallest (w.r.t set $\subseteq$) complete extension.*

The status of arguments is defined in terms of the acceptability semantics.

Definition 6 (Argument status). *Let $T = \langle \mathcal{A}, \texttt{Def} \rangle$ be a argumentation system, and $\mathcal{E}_1, \ldots, \mathcal{E}_x$ its extensions under a given semantics. Let $a \in \mathcal{A}$.*

- *a is* skeptically accepted *in T iff $a \in \mathcal{E}_{i=1,\ldots,x}$, $\forall \mathcal{E}_i \neq \emptyset$.*
- *a is* credulously accepted *in T iff $\exists \mathcal{E}_i$ such that $a \in \mathcal{E}_i$.*
- *a is* rejected *in T iff $\nexists \mathcal{E}_i$ such that $a \in \mathcal{E}_i$.*

A argumentation systems favors offers (or options) that are supported by the either credulous or skeptical arguments. These offers are the ”outcomes” of the system.

Definition 7 (Outcome of argumentation system). *The* outcome *of a argumentation system* $T = (\mathcal{O}, \mathcal{A}, \texttt{Def})$ *is* $\texttt{Outcome}(T) = \{o_i \in \mathcal{O}$ *such that* $\exists a \in \mathcal{F}(o_i)$ *and* a *is skeptically or credoulously accepted*$\}$.

Since the outcome of the system is determined solely by their supporting arguments, we can restrict our analysis to the properties of the arguments of the system.

3 Aggregating Argumentation Systems

Throughout this section, we assume two agents with the argumentation systems $T_1 = \langle \mathcal{O}, \mathcal{A}, \succeq_1, \texttt{Def}_1 \rangle$ and $T_2 = \langle \mathcal{O}, \mathcal{A}, \succeq_2, \texttt{Def}_2 \rangle$, respectively. We capture the aggregation of two such argumentation systems in a new argumentation system $U^{T_1,T_2} = \langle \mathcal{O}, \mathcal{A}, \succeq, \texttt{Def} \rangle$, whose defeat relation is the *union* of the defeat relations of the individual systems.

Definition 8 (Aggregate Argumentation System). *Let* $T_1 = \langle \mathcal{O}, \mathcal{A}, \succeq_1, \texttt{Def}_1 \rangle$ *and* $T_2 = \langle \mathcal{O}, \mathcal{A}, \succeq_2, \texttt{Def}_2 \rangle$ *be the argumentation systems of two agents. The* aggregate argumentation system *or simply* aggregate system *of* T_1, T_2 *is* $U^{T_1,T_2} = \langle \mathcal{O}, \mathcal{A}, \succeq, \texttt{Def}_1 \cup \texttt{Def}_2 \rangle$.

To complete the picture of the aggregate argumentation system, we characterize the preference relation $\succeq$, underlying the defeat relation $\texttt{Def}_1 \cup \texttt{Def}_2$, in terms of the individual preference relations $\succeq_1$ and $\succeq_2$. It turns out that the union of the defeat relations of the individual systems corresponds exactly to the defeat relation induced by the intersection (i.e. common preferences) of the individual preference relations $\succeq_1$ and $\succeq_2$. We believe that this a suitable way to take into account the interests of the negotiating agents, which is a main point in the integrative negotiation, and corresponds to a cooperative approach to mutual problem solving. In order to present compactly this result, the following definition is needed.

Definition 9. *Let* $\succeq$ *be a preference relation on* A. *We denote by* $\mathit{Def}(\succeq)$ *the* Def *relation induced by* $\succeq$ *and the weakly complete relation on* A.

Theorem 1. *Let* $T_1 = \langle \mathcal{A}, \texttt{Def}_1 \rangle$ *and* $T_2 = \langle \mathcal{A}, \texttt{Def}_2 \rangle$ *be argumentation systems, where* $\texttt{Def}_1$ *and* $\texttt{Def}_2$ *are defined on the preference relations* $\succeq_1$ *and* $\succeq_2$ *respectively. Then* $\texttt{Def}_1 \cup \texttt{Def}_2 = \mathit{Def}(\succeq_1 \cap \succeq_2)$.

Consequently, the union aggregation of theories T_1 and T_2 above is $U^{T_1,T_2} = \langle \mathcal{O}, \mathcal{A}, \succeq_1 \cap \succeq_2, \texttt{Def}_1 \cup \texttt{Def}_2 \rangle$. Note that the sets of options $\mathcal{O}$ is the same for both agents and the aggregate system. Following our convention, we denote omit $\mathcal{O}$ and the preference relations, and denote the individual theories by $T_1 = \langle \mathcal{A}, \texttt{Def}_1 \rangle$, $T_2 = \langle \mathcal{A}, \texttt{Def}_2 \rangle$, and their aggregate theory by $U^{T_1,T_2} = \langle \mathcal{A}, \texttt{Def}_1 \cup \texttt{Def}_2 \rangle$.

The following example illustrates the features of the union aggregation.

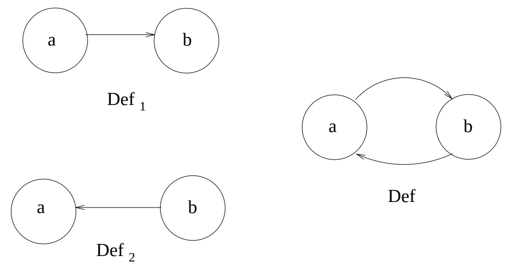

Figure 1. Individual theories and aggregate theory of agents with contradictory preferences, where $\texttt{Def} = \texttt{Def}_1 \cup \texttt{Def}_2$

.

Example 1. *Let* $\mathcal{A} = \{a, b\}$ *be a set of arguments common to agents* α *and* β*, having the defeat relations* $\texttt{Def}_1 = \{(a, b)\}$ *and* $\texttt{Def}_2 = \{(b, a)\}$ *respectively. Following Definition* 2 *and Property* 2*, we conclude that for agent* α *it must be the case that* $a \succ b$*, whereas for agent* β *it holds that* $b \succ a$*. We therefore observe that the two agents have conflicting preferences, as* α *strictly prefers* a *to* b*, whereas* β *strictly prefers* b *to* a*. By applying* union *to* $\texttt{Def}_1$ *and* $\texttt{Def}_2$*, we obtain the relation* $\texttt{Def}_1 \cup \texttt{Def}_2 = \{(a, b), (b, a)\}$*. The situation is depicted graphically in figure 1. Following again Definition* 2 *and Property* 2*, we conclude that arguments* a *and* b *are now incomparable for agents* α *and* β*, if we regard them as a group. This means that these agents will be unable to express a common preference on arguments* a *and* b*, and this captures perfectly the current situation, namely that the two agents have contradictory preferences on these arguments.*

Consider the argumentation systems of agents α *and* β*, on the set of arguments* $\mathcal{A} = \{a, b, c\}$*, and defeat relations* $\texttt{Def}_1 = \{(a, c), (b, c)\}$ *and* $\texttt{Def}_2 = \{(a, c), (b, c)\}$ *respectively. It must be the case that for both agents,* $a \succ c$ *and* $b \succ c$*. We also conclude that for both agents,* $a \succeq b$ *and* $b \succeq a$*, because* $(a, b), (b, a) \notin \texttt{Def}_1$ *and* $(a, b), (b, a) \notin \texttt{Def}_2$*. In other words, agents* α *and* β *strictly prefer* a *to* c *and* b *to* c *and are indifferent between* a *and* b*. It is easy to see that this is exactly the result we obtain by the union aggregation of* $\texttt{Def}_1$ *and* $\texttt{Def}_2$*, namely* $\texttt{Def}_1 \cup \texttt{Def}_2 = \{(a, c), (b, c)\}$.

4 Properties of Aggregate Systems

In this section we present the formal properties of the aggregate system defined in the previous section.

Although the defeat relation of the aggregate system needs not to be transitive in general, there are cases where subparts of this relation have this property. One such case is that of *unidirectional chains* that are defined formally below. As it will become apparent in the following, transitivity imposes a strong structure with significant implications. This is an important difference between the defeat relations that are investigated in this work and the attacking relations of the abstract framework of Dung, that lack specific structure.

Definition 10. *A* unidirectional chain *of a binary relation R on a set A is a sequence $a_1, a_2, \ldots, a_n$ of elements of A such that $(a_i, a_{i+1}) \in R$ and $(a_{i+1}, a_i) \notin R$, for $1 \leq i \leq n$.*

Based on the above definition, we obtain the following result.

Theorem 2. *Every unidirectional chain of $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$ is transitive.*

The next theorem shows that the admissible arguments coincide with the *self-defending* ones. The notion of *self-defending argument* is formally defined as follows.

Definition 11. *Let $T = \langle \mathcal{A}, \mathtt{Def} \rangle$ be a argumentation system and let $a \in \mathcal{A}$. Then a is* self-defending *in T if $\forall a' \in \mathcal{A}$ such that $(a', a) \in \mathtt{Def}$, it holds that $(a, a') \in \mathtt{Def}$.*

Formally, the correspondence between admissible and self-defending arguments is as follows.

Theorem 3. *Let $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ and $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ be argumentation systems, $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$, $\mathcal{E}$ a preferred extension of U^{T_1,T_2}, and $a \in \mathcal{A}$. Then $a \in \mathcal{E}$ iff a is self-defending in U^{T_1,T_2}.*

The following theorem shows that aggregate argumentation systems are *coherent* (i.e. each preferred extension is also stable), which is an important property for argumentation systems.

Theorem 4. *Let $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ and $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ be argumentation systems and $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$. Every preferred extension of U^{T_1,T_2} is a stable extension of U^{T_1,T_2}.*

Similarly to the credulously accepted arguments of an aggregate system, the skeptically accepted ones can be characterized concisely by their properties. Indeed, these arguments are exactly those that have no defeaters. These arguments have in-degree 0 in the graph associated with this system. (Recall that the in-degree of a node in a directed graph is the number of arcs that have this node as sink in the graph). If a is argument of an argumentation system T, its in-degree in T is denoted by $in_T(a)$.

Theorem 5. *Let $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ and $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ be argumentation systems, $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$, and $a \in \mathcal{A}$ a skeptically accepted argument of U^{T_1,T_2}. Then $in_{U^{T_1,T_2}}(a) = 0$.*

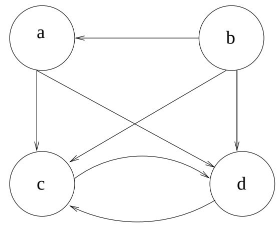

Figure 2. Aggregate Agent Theory.

The following example recapitulates the previous analysis.

Example 2. *Let $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$ be an aggregate argumentation system, with $\mathcal{A} = \{a, b, c, d\}$, $\mathtt{Def_1} = \{(a, c), (a, d), (b, c), (b, d), (c, d)\}$ and $\mathtt{Def_2} = \{(a, c), (a, d), (b, a), (b, c), (b, d), (c, d)\}$. System U^{T_1,T_2} is presented in figure 2. The only node that is self-defending is b, so U^{T_1,T_2} has a stable extension that contains b. Moreover, $in(b) = 0$, and therefore b is also a skeptical conclusion. If the element (b, a) is removed from $\mathtt{Def_2}$, both nodes a and b become self-defending in system U^{T_1,T_2}, and both of them have in-degree 0. In fact, U^{T_1,T_2} has one stable extension that contains both, therefore they are both skeptically accepted.*

Finally assume the initial $\mathtt{Def_2}$ relation, and suppose that (a, b) is added to $\mathtt{Def_1}$. Then, again both of a and b are self-defending in U^{T_1,T_2}, therefore they must belong to some extension of the theory. Furthermore, none of them has now in-degree 0, so none is skeptically accepted. Indeed, U^{T_1,T_2} has two different extensions, namely $\mathcal{E}_1 = \{a\}$ and $\mathcal{E}_2 = \{b\}$, and no skeptically accepted arguments.

It is important to notice the computational implications of the properties of the credulously and skeptically accepted arguments. Checking whether an argument a is a credulous conclusion of an aggregate system T is equivalent to checking whether it is self-defending, a computation that can be performed in polynomial time. Similarly, to check whether an argument is a skeptical conclusion, it suffices to verify that it has no defeaters, a computation that is also polynomial. Contrast, this polynomial time algorithms with the intractability results known for general argumentation frameworks.

Another property of the extensions of aggregate argumentation systems is that every member of such an extension is incomparable to any other element of any other such extension. Formally this is stated as follows.

Theorem 6. *Let $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ and $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ be argumentation systems, $\mathcal{E}_1, \ldots \mathcal{E}_k$ be the extensions of the system $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$, and $a, b \in \mathcal{A}$ be two arguments such that $a, b \in \cup_{i=1}^{k} \mathcal{E}_i$ and $\neg \exists \mathcal{E}_i$ such that*

$a, b \in \mathcal{E}_i$*, for* $1 \leq i \leq k$*. Then* $(a, b) \in \mathtt{Def_1} \cup \mathtt{Def_2}$ *and* $(b, a) \in \mathtt{Def_1} \cup \mathtt{Def_2}$.

To complete the picture of the structures arising in aggregate argumentation systems, we prove that the stable extensions of these systems are disjoint. Combined with the previous result, this basically means that incomparability partitions the set of self-defending arguments into different extensions.

Theorem 7. *If* $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ *and* $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ *are argumentation systems, the stable extensions of the system* $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$ *are disjoint.*

From the above it follows immediately that U^{T_1,T_2} has a skeptical conclusion iff it has exactly one stable extension.

Theorem 8. *Let* $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ *and* $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ *be argumentation systems, and* $\mathcal{E}$ *the unique stable extension of* $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$*. Then,* $\mathcal{E}$ *is a grounded extension of* U^{T_1,T_2}.

Our analysis to this point has revealed some of the properties enjoyed by the aggregate argumentation systems. Moreover, these properties form the basis for studying the relation between the individual theories and the aggregate system constructed by taking the union of the defeat relations.

The first result on the relation between the individual and the aggregate argumentation systems, states that the set of skeptically accepted arguments of the individual systems coincides with the skeptically accepted arguments of the aggregate system.

Theorem 9. *Let* $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ *and* $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ *be argumentation systems and* $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$ *with stable extensions* $\mathcal{E}_i^{T_1}$*,* $\mathcal{E}_j^{T_2}$*,* $\mathcal{E}_k^{U^{T_1,T_2}}$ *respectively, and* $1 \leq i \leq m$*,* $1 \leq j \leq p$*,* $1 \leq k \leq q$*. Then* $\cap_{i=1}^{m} \mathcal{E}_i^{T_1} \bigcap \cap_{j=1}^{p} \mathcal{E}_j^{T_2} = \cap_{k=1}^{q} \mathcal{E}_k^{U^{T_1,T_2}}$.

The relation between the credulous conclusions of the individual and aggregate argumentation systems is more complicated. One relation that holds between the two sets is that the common credulous conclusions of the individual theories are also credulous conclusions of the aggregate system.

Theorem 10. *Let* $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ *and* $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ *be argumentation systems and* $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$ *with preferred extensions* $\mathcal{E}_i^{T_1}$*,* $\mathcal{E}_j^{T_2}$*,* $\mathcal{E}_k^{U^{T_1,T_2}}$ *respectively, and* $1 \leq i \leq m$*,* $1 \leq j \leq p$*,* $1 \leq k \leq q$*. Then* $\cup_{i=1}^{m} \mathcal{E}_i^{T_1} \bigcap \cup_{j=1}^{p} \mathcal{E}_j^{T_2} \subseteq \cup_{k=1}^{q} \mathcal{E}_k^{U^{T_1,T_2}}$.

The observations that are presented in the following reveal that the link between the credulous conclusions of the individual and the aggregate argumentation systems is not as strong as the one between skeptical conclusions. First observe that the converse of the above theorem does not hold, i.e. $\cup_{k=1}^{q} \mathcal{E}_k^{U^{T_1,T_2}} \subseteq \cup_{i=1}^{m} \mathcal{E}_i^{T_1} \bigcap \cup_{j=1}^{p} \mathcal{E}_j^{T_2}$ is not necessarily true. Assume $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ and $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$, and let $\mathcal{A} = \{a, b\}$, $\mathtt{Def_1} = \{(a, b)\}$, $\mathtt{Def_2} = \{(b, a)\}$. Then, U^{T_1,T_2} has two stable extensions with $\cup_{k=1}^{2} \mathcal{E}_k^{U^{T_1,T_2}} = \{a, b\}$, whereas $\mathcal{E}^{T_1} \cap \mathcal{E}^{T_2} = \emptyset$.

Moreover, observe that the relation $\cup_{i=1}^{m} \mathcal{E}_i^{T_1} \bigcup \cup_{j=1}^{p} \mathcal{E}_j^{T_2} \subseteq \cup_{k=1}^{q} \mathcal{E}_k^{U^{T_1,T_2}}$ does not necessarily hold. Consider for instance the argumentation systems T_1, T_2 built from $\mathcal{A} = \{a, b\}$ and the relations $\mathtt{Def_1} = \emptyset$, and $\mathtt{Def_2} = \{(a, b)\}$. Then, $\cup_{i=1}^{m} \mathcal{E}_i^{T_1} \bigcup \cup_{j=1}^{p} \mathcal{E}_j^{T_2} = \{a, b\}$, whereas $\cup_{k=1}^{q} \mathcal{E}_k^{U^{T_1,T_2}} = \{a\}$.

More interestingly, it is not necessarily true that $\cup_{k=1}^{q} \mathcal{E}_k^{U^{T_1,T_2}} \subseteq \cup_{i=1}^{m} \mathcal{E}_i^{T_1} \bigcup \cup_{j=1}^{p} \mathcal{E}_j^{T_2}$. To see this consider the argumentation systems T_1, T_2 built from $\mathcal{A} = \{a, b, c\}$ and the relations $\mathtt{Def_1} = \{(c, a), (c, b), (a, b)\}$, and $\mathtt{Def_2} = \{(b, a), (b, c), (a, c)\}$. Then, U^{T_1,T_2} has three stable extensions with $\cup_{k=1}^{3} \mathcal{E}_k^{U^{T_1,T_2}} = \{a, b, c\}$, whereas $\mathcal{E}^{T_1} \cup \mathcal{E}^{T_2} = \{b, c\}$. Note that $\{a\}$ is an extension of U^{T_1,T_2}, despite the fact that a does not belong to any of the stable extensions of the systems of the individual agents.

This discrepancy between the conclusions drawn from the individual theories and the aggregate theory becomes more interesting in the light of the following results. It states that the credulous conclusions of the aggregate theory are in a direct correspondence with the *Pareto optimal* arguments of the theory.

A general definition of Pareto optimality in the context of argumentation making with multiple criteria, can be given through the notion of *dominance*. Briefly speaking, a solution *dominates* another solution if it is better or equally good in all criteria and strictly better in at least one criterion. A solution S is called *Pareto optimal*, if there is no other solution that dominates S. In our framework, Pareto optimality is defined as follows.

Definition 12. *Let* $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ *and* $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ *be argumentation systems, where* $\mathtt{Def_1}$ *and* $\mathtt{Def_2}$ *are defined on the preference relations* $\succeq_1$ *and* $\succeq_2$ *respectively. An arguments* $a \in \mathcal{A}$ *is Pareto optimal wrt* T_1 *and* T_2 *if there is no other argument* $a' \in \mathcal{A}$ *such that* $a' \succeq_1 a$ *and* $a' \succ_2 a$ *or* $a' \succeq_2 a$ *and* $a' \succ_1 a$*.*

Theorem 11. *Let* $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ *and* $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ *be argumentation systems and* $U^{T_1,T_2} = \langle \mathcal{A}, \mathtt{Def_1} \cup \mathtt{Def_2} \rangle$ *with preferred extensions* $\mathcal{E}_i^{U^{T_1,T_2}}$ *with* $1 \leq i \leq k$*. Then* $a \in \cup_{i=1}^{k} \mathcal{E}_i^{U^{T_1,T_2}}$ *iff* a *is a Pareto optimal argument wrt* T_1 *and* T_2*.*

The following example illustrates the notion of Pareto optimality in the context of argumentation systems.

Example 3. *Let* $T_1 = \langle \mathcal{A}, \mathtt{Def_1} \rangle$ *and* $T_2 = \langle \mathcal{A}, \mathtt{Def_2} \rangle$ *be argumentation systems, where* $\mathcal{A} = \{a, b, c, d\}$ *and* $\mathtt{Def_1}$*,* $\mathtt{Def_2}$*, defined on the underlying pre-orders* $\succeq_1$ *and* $\succeq_2$*,* *which are the following:*

$a \succeq_1 b \quad a \succeq_1 c \quad a \succeq_1 d \quad b \succeq_1 c \quad b \succeq_1 d \quad c \succeq_1 d$
$d \succeq_2 c \quad d \succeq_2 a \quad d \succeq_2 b \quad c \succeq_2 a \quad c \succeq_2 b \quad a \succeq_2 b$
The argumentation systems T_1 and T_2 have the extensions $\mathcal{E}^{T_1} = \{a\}$ and $\mathcal{E}^{T_2} = \{d\}$, respectively. The extension of their aggregation is $\mathcal{E}^{U^{T_1,T_2}} = \{a, d, c\}$. Note that the elements of $\mathcal{E}^{U^{T_1,T_2}}$ are Pareto optimal arguments, and $c \in \mathcal{E}^{U^{T_1,T_2}} - (\mathcal{E}^{T_1} \cup \mathcal{E}^{T_2})$. To explain the intuitive meaning of Pareto optimality in the context of argumentation systems, we consider argument a, and assume that both agents agree on a (and therefore on the outcome supported by a). Then, there is no other argument in $\mathcal{A}$ on which they could both agree and both *improve wrt their preferences. For instance, if they both abandon a in favor of d, then the agent with the preference relation $\succeq_2$ improves, but the agent with $\succeq_2$ is worse off. Finally, argument b is not Pareto optimal, since instead of agreeing on b, both agents can agree on a and thereof* both *improve their preferences, given that $a \succeq_1 b$ and $a \succeq_2 b$.*

5 Conclusions and Related Work

This paper proposed a characterization of the outcomes of argumentation-based integrative negotiations. For this reason we considered the ”ideal” situation where agents have complete information or negotiate through a mediator. In order to capture this situation we proposed the aggregation of individual negotiation theories, represented as preference-based argumentation theories, and we investigated the properties of the *union* aggregation of such argumentation systems. In a nutshell, this method merges together the individual argumentation systems by taking as its underlying preference relation the intersection of the individual preference relations. Our study of this aggregation operator revealed that it has many desirable theoretical and computational features and therefore it is very suitable for capturing the essence of the integrative negotiation where the key issue is to look for solutions that are ”good” for all agents.

Among the presented results, we want to emphasize the fact that the outcomes of the aggregate theory correspond to Pareto optimal solutions wrt the preferences of the individual systems. We also showed that these outcomes, under standard semantics for argumentation, have strong structural properties. For instance, the arguments that belong to the extensions are self-defending, while the skeptical conclusions are exactly those arguments that have no defeaters. These properties translate directly to polynomial time algorithms for credulous and skeptical reasoning. This could be very useful in the situations where the negotiation is made through *mediator* who receives the arguments of the negotiating parts and tries to find a compromise for them. In the current paper we considered the case of bilateral negotiations by aggregating two agents’ theories. The same results apply in the case of several agents as well.

Integrative negotiation has already been studied in the context of multi-agent systems (see e.g. [12, 8]). However, to our knowledge it is the first time that a theoretical investigation of the characteristics of the outcomes of an argumentation-based integrative negotiation framework has been done.

The outcomes found on the basis of the aggregate theory and corresponding to the ”ideal” situation, can also be used as a baseline for the evaluation of different integrative negotiation protocols and strategies in the cases where the negotiation takes place in a distributed manner. This is an issue for our future work. Moreover, we will investigate the implications of the ideas presented in [4, 11] for the aggregation of classical Dung’s argumentation systems.

References

[1] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–216, 2002.

[2] L. Amgoud, Y. Dimopoulos, and P. Moraitis. A unified and general framework for argumentation-based negotiation. In *Proc. AAMAS’07*, pages 963–970, 2007.

[3] L. Amgoud, Y. Dimopoulos, and P. Moraitis. Making decisions through preference-based argumentation. In *Proc. KR’08*, 2008.

[4] S. Coste-Marquis, C. Devred, S. Konieczny, M. Lagasquie-Schiex, and P. Marquis. Merging argumentation systems. In *Proc. AAAI’05*, pages 614–619, 2005.

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

[6] R. Fisher and W. Ury. *Getting to Yes: Negotiating Agreement Without Giving In,*. Penguin Books, 1991.

[7] N. R. Jennings, P. Faratin, A. R. Lumuscio, S. Parsons, and C. Sierra. *Automated negotiation: Prospects, methods and challenges*. Group Decision and Negotiation, 2001.

[8] P. Pasquier, L. Sonenberg, I. Rahwan, F. Dignum, and R. Hollands. An empirical study of interest-based negotiation. In *Proc.ICEC’07*, pages 339–348, 2007.

[9] I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and E. Sonenberg. Argumentation-based negotiation. *Knowledge Engineering Review*, 18 (4):343–375, 2003.

[10] J. Rosenschein and G. Zlotkin. *Rules of Encounter: Designing Conventions for Automated Negotiation Among Computers,*. MIT Press, Cambridge, Massachusetts, 1994., 1994.

[11] F. Tohm, G. Bodanza, and G. Simari. Aggregation of attack relations: A social-choice theoretical analysis of defeasibility criteria. In *FoIKS*, pages 8–23, 2008.

[12] X. Zhang, V. Lesser, and T. Wagner. Integrative Negotiation Among Agents Situated in Organizations. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 36(1):19–30, January 2006.