Generation and evaluation of different types of arguments in negotiation

Leila Amgoud and **Henri Prade**

Institut de Recherche en Informatique de Toulouse (IRIT)
118, route de Narbonne, 31062 Toulouse, France
{amgoud, prade}@irit.fr

Abstract

Until now, AI argumentation-based systems have been mainly developed for handling inconsistency. In that explanation-oriented perspective, only one type of argument has been considered. Several argumentation frameworks have then been proposed for generating and evaluating such arguments. However, recent works on argumentation-based negotiation have emphasized different other types of arguments such as *threats*, *rewards*, *appeals*, etc...
The purpose of this paper is to provide a logical framework which encompass the classical argumentation-based framework and handles the new types of arguments. More precisely, we give the logical definitions of these arguments and their weighting systems. These definitions take into account that negotiation dialogues involve not only agents' beliefs (of various strengths) but also their goals (having maybe different priorities), the beliefs on the goals of other agents, etc... In other words, from the different belief and goal bases maintained by an agent, we can generate all the possible threats, rewards, explanations, appeals which are associated to them. Finally, we show how to evaluate conflicting arguments of different types. The possibilistic logic framework is used for handling formulas with different degrees of certainty or priority.

Key words: Negotiation, Argumentation

Introduction

Argumentation is a promising approach for reasoning with inconsistent knowledge, based on the construction and the comparison of arguments. It may be also considered as a different method for handling uncertainty. A basic idea behind argumentation is that it should be possible to say more about the certainty of a particular fact than assessing a certainty degree in $[0, 1]$. In particular, it should be possible to assess the reason why a fact holds, under the form of arguments, and combine these arguments for the certainty evaluation. Indeed, the process of combination may be viewed as a kind of reasoning about the arguments in order to determine the most acceptable of them. Various argument-based frameworks to defeasible reasoning have been developed (Amgoud & Cayrol 2002a; 2002b; Dung 1995; Prakken & Sartor 1997) for generating and evaluating arguments. In that explanation-oriented perspective, only one type of argument has been considered. Namely, what we call *explanatory* arguments.

Recent works on negotiation (Amgoud & Prade 2004; Kraus, Sycara, & Evenchik 1998; Parsons, Sierra, & Jennings 1998; Rahwan *et al.* 2004; Ramchurn, Jennings, & Sierra 2003) have argued that argumentation plays a key role in finding a compromise. Indeed, an offer supported by a 'good argument' has a better chance to be accepted by another agent. Argumentation may also lead an agent to change its goals and finally may constrain an agent to respond in a particular way. For example, if an agent receives a threat, this agent may accept the offer even if it is not really acceptable for it.
In addition to explanatory arguments studied in classical argumentation frameworks, the above works on argumentation-based negotiation have emphasized different other types of arguments such as *threats*, *rewards*, *appeals*, etc... In (Kraus, Sycara, & Evenchik 1998; Ramchurn, Jennings, & Sierra 2003), these arguments are treated as speech acts with pre-conditions and post-conditions.

The purpose of this paper is to provide a logical framework which encompasses the classical argumentation-based framework and handles the new types of arguments. More precisely, we give the logical definitions of these arguments and their weighting systems. These definitions take into account the fact that negotiation dialogues involve not only agents' beliefs (of various strengths), but also their goals (having maybe different priorities), and the beliefs on the goals of other agents. Thus, from the different belief and goal bases maintained by an agent, we can generate all the possible threats, rewards, explanations, appeals, which are associated to them. Finally, we show how to evaluate conflicting arguments of different types. The possibilistic logic framework is used for handling formulas with different degrees of certainty or priority. An illustrative example is grounded in the last section.

Types of arguments

In what follows, $\mathcal{L}$ will denote a propositional language. $\vdash$ denotes classical inference and $\equiv$ denotes logical equivalence.
We suppose that we have two negotiating agents: P (called also a proponent) and C (called also an opponent). In all what follows, we suppose also that P presents an argument

to C. Each negotiating agent is supposed to have a set $\mathcal{G}$ of *goals* to pursue, a knowledge base, $\mathcal{K}$, gathering the information it has about the environment, and finally a base $\mathcal{GO}$, containing what the agent believes the goals of the other agent are, as already assumed in (Amgoud & Prade 2004).
$\mathcal{K}$ may be pervaded with uncertainty (the beliefs are more or less certain), and the goals in $\mathcal{G}$ and $\mathcal{GO}$ may not have equal priority. Thus, levels of certainty are assigned to formulas in $\mathcal{K}$, and levels of priority are assigned to the goals. We obtain three possibilistic bases (Dubois, Lang, & Prade 1991) that model gradual knowledge and preferences:

$$\mathcal{K} = \{(k_i, \alpha_i), i = 1, \ldots, n\},$$
$$\mathcal{G} = \{(g_j, \beta_j), j = 1, \ldots, m\},$$
$$\mathcal{GO} = \{(go_l, \delta_l), l = 1, \ldots, p\}$$

where k_i, g_j, go_l are propositions of the language $\mathcal{L}$ and α_i, β_j, δ_l are elements of $[0, 1]$, or of any linearly ordered scale, finite or not.
We shall denote by $\mathcal{K}^*$, $\mathcal{G}^*$ and $\mathcal{GO}^*$ the corresponding sets of classical propositions when weights are ignored, i.e.

$$\mathcal{K}^* = \{k_i, i = 1, \ldots, n\},$$
$$\mathcal{G}^* = \{g_j, j = 1, \ldots, m\},$$
$$\mathcal{GO}^* = \{go_l, l = 1, \ldots, p\}$$

We distinguish between three categories of arguments according to their logical definitions: *the threats, the rewards* and *the explanatory arguments*. In what follows we will discuss each category of arguments.

Threats

Threats are very common in human negotiation. They have a negative character and are applied to force an agent to behave in a certain way. Two forms of threats can be distinguished:

- You should do α otherwise I will do β
- You shouldn't do α otherwise I will do β

The first case occurs when an agent P needs an agent C to do α and C refuses. P threatens then C by doing β which, according to its beliefs, will have bad consequences for C. Let's consider the following example:

Example 1 *A mother asks her child to carry out his school work and he refuses. The mother then threatens him not to let him go to the festival organized by her friend the next week-end.*

The second kind of threats occurs when an agent C wants to do some action α, which is not acceptable for P. In this case, P threatens that if C insists to do α then it will do β which, according to P's beliefs, will have bad consequences for C. To illustrate this kind of threat, we consider the following example borrowed from (Kraus, Sycara, & Evenchik 1998).

Example 2 *A labor union insists on a wage increase. The management says it cannot affort it, and asks the union to withdraw its request. The management threatens that, if it grants this increase, it will have to lay off employees to compensate for the higher operational cost that the increase will entail.*

In fact, for a threat to be effective, it should be paintful for its receiver and conflicts at least one of its goals. A threat is then made up of three parts: i) the conclusion that the agent who makes the threat wants, the threat itself and finally the threatened goal. IN the case of of example 1, the mother has a threat in favour of doing the school work. Formally:

Definition 1 (Threat) *A* threat *is a triple* $<H, h, \phi>$ *such that:*

1. $H \subseteq \mathcal{K}^*$,
2. $H \cup \{\neg h\} \vdash \neg\phi$ *such that* $\phi \in \mathcal{GO}^*$,
3. $H \cup \{\neg h\}$ *is consistent and minimal (for set inclusion) among the sets satisfying the two first conditions.*

$\mathcal{A}_t$ *will denote the set of all threats that may be constructed from the bases* $<\mathcal{K}, \mathcal{G}, \mathcal{GO}>$. H *is the* support *of the threat,* h *its* conclusion *and* ϕ *is the* threatened goal.

Note that the above definition captures the two forms of threats.

Example 3 *Let's consider an agent* P *having the three following bases:* $\mathcal{K} = \{(\neg\, finish - work \rightarrow overtime, 1)\}$, $\mathcal{G} = \{(finish - work, 1)\}$ *and* $\mathcal{GO} = \{(\neg overtime, 0.7)\}$. *Let's suppose that the agent* P *asks the agent* C *to finish the work and that C refuses.* P *can then make the following threat:* $< \{\neg\, finish{-}work \rightarrow overtime\}, finish{-}work, \neg overtime>$.

Rewards

During a negotiation an agent P can entice agent C to do α by offering to do an action β as a reward. Of course, agent P believes that β will contribute to the goals of C. Thus, a reward has generally, at least from the point of view of its sender, a positive character. As for threats, two forms of rewards can be distinguished: "If you do α then I will do β" and "If you don't do α then I will do β".

Example 4 *A sales agent tries to persuade a customer to buy a computer by offering a set of blank cassettes.*

Formally, a reward is defined as follows:

Definition 2 (Reward) *A* reward *is a triple* $<H, h, \phi>$ *such that:*

1. $H \subseteq \mathcal{K}^*$,
2. $H \cup \{h\} \vdash \phi$ *such that* $\phi \in \mathcal{GO}^*$,
3. $H \cup \{h\}$ *is consistent and minimal (for set inclusion) among the sets satisfying the two first conditions.*

$\mathcal{A}_r$ *will denote the set of all the rewards that can be constructed from* $<\mathcal{K}, \mathcal{G}, \mathcal{GO}>$. H *is the* support *of the reward,* h *its* conclusion *and* ϕ *the* rewarded goal.

Example 5 *Let's consider an agent* P *having the three following bases:* $\mathcal{K} = \{(finish - work \rightarrow high - budget, 1), (high - budget \rightarrow high - salary, 0.6)\}$, $\mathcal{G} = \{(finish - work, 1)\}$ *and* $\mathcal{GO} = \{(high - salary, 1)\}$.
We suppose the agent P *asks* C *to finish the work and* C *refuses.* P *can then present the following reward in favour of its offer/ request "finish-work":* $<\{finish{-}work \rightarrow high - budget,\ high - budget \rightarrow high - salary\}, finish - work, high - salary>$.

In (Kraus, Sycara, & Evenchik 1998), another kind of arguments has been pointed out. It is the so-called *appeal to self-interest*. In this case, an agent P believes that the suggested offer implies one of C's goals. In fact, this case may be seen as a *self-reward* and consequently it is a particular case of rewards.

Explanatory arguments

Explanations constitute the most common category of arguments. In classical argumentation-based frameworks which have been developed for handling inconsistency in knowledge bases, each conclusion is justified by arguments. They represent the reasons to believe in the fact. Such arguments have a deductive form. Indeed, from premisses, a fact or a goal is entailed. Formally:

Definition 3 (Explanatory argument) *An* explanatory *argument is a pair <H, h> such that: i) $H \subseteq \mathcal{K}^* \cup \mathcal{G}^* \cup \mathcal{GO}^*$, ii) $H \vdash h$, iii) H is consistent and minimal (for set inclusion).*
$\mathcal{A}_e$ will denote the set of all the explanatory arguments that can be constructed from $<\mathcal{K}, \mathcal{G}, \mathcal{GO}>$. H is the support *of the argument and h its* conclusion.

Example 6 *Let's consider the case of an agent who wants to go to Sydney. $\mathcal{K}$ = $\{(conference, 0.8), (cancelled, 0.4), (conference \rightarrow Sydney, 1), (cancelled \rightarrow \neg conference, 1)\}$. $\mathcal{G}$ = $\{(Sydney, 1)\}$ and $\mathcal{GO}$ = $\emptyset$.*

The agent wants to go to Sydney and justifies his wish by the following explanatory argument: $<\{conference, conference \rightarrow Sydney\}, Sydney>$. Indeed, from its beliefs one can deduce the fact $Sydney$.

In (Kraus, Sycara, & Evenchik 1998), other kinds of arguments have been proposed and they are called *appeals*. We argue that the different forms of appeals can be modelled as explanatory arguments. In what follows, we will show through examples how such appeals are defined as explanatory arguments.

An appeal to prevailing practice: In this case, the agent believes that the opponent agent refuses to perform the requested action since it contradicts one of its own goals. However, the agent gives a counter-example from a third agent's actions, hoping it will serve as a convincing evidence. Of course, the third agent should have the same goals as the opponent and sould have performed the action successfully.

Example 7 *An agent P asks another agent C to make overtime. C refuses because it is afraid that this is punished by law. The bases of C are then: $\mathcal{K}$ = $\{(overtime \rightarrow ToBePunished, 1)\}$, $\mathcal{G}$ = $\{(\neg ToBePunished, 1)\}$ and $\mathcal{GO}$ = $\emptyset$.*
When the opponent C receives the offer $overtime$, it constructs an explanatory argument in favor of $ToBePunished$: $<\{overtime, overtime \rightarrow ToBePunished\}, ToBePunished>$. This argument confirms to him that its goal will be violated then it refuses the offer. The proponent P reassures him by telling that another collegue makes overtime and it never has problems with law. In fact, it presents the following counter-argument: $<\{overtime, \neg ToBePunished\}, \neg(overtime \rightarrow ToBePunished)>$. This last argument is an appeal to prevailing practice.

An appeal to past promise: In this case, the agent expects the opponent agent to perform an action based on past promise. Let's illustrate it by the following example:

Example 8 *A child asks his mother to buy a gift to him and the mother refuses. The child points out that she promised to buy something to him if he succeeds at his examinations. The bases of the child are: $\mathcal{K}$ = $\{$(success, 1), $(success \rightarrow gift, 1)\}$, $\mathcal{G}$ = $\{(gift, 1)\}$ and $\mathcal{GO}$ = $\emptyset$.*
The child's argument is then: $<\{success, success \rightarrow gift\}, gift>$.

A counter-example: This argument is similar to "appeal to prevailing practice"; however, the counter-example is taken from the opponent agent's own history of activities. In this case, the counter argument produced by the proponent should be constructed from the beliefs of the opponent. In the case of example 7, the support of the counter-argument should be included in the base of C. Thus, C would have a conflicting base.
These three types of arguments have the same nature and they are all deductive. They are defined logically as explanatory arguments. The nature of these arguments, however, plays a key role in the strategies used by the agents. For example, a counter-example may lead quickly the other agent to change its mind than an appeal to prevailing practice.
In what follows, we denote by $\mathcal{A}_x$ the set of arguments of nature x with $x \in \{t, r, e\}$.

The strengths of the arguments

In (Amgoud & Cayrol 2002a), it has been argued that arguments may have different forces according to the beliefs from which they are constructed. The basic idea is that arguments using more certain beliefs are stronger than arguments using less certain beliefs. Thus, a level of certainty is assigned to each argument. These certainty levels make it possible to compare arguments. In fact, an argument A is preferred to another argument B iff A is stronger than B.
As mentioned before, each of the three bases $<\mathcal{K}, \mathcal{G}, \mathcal{GO}>$ is pervaded with uncertainty or equipped with priority levels. From these degrees, we first define the force of an explanatory argument.

Definition 4 (Force of an explanatory argument) *Let A = $<H, h> \in \mathcal{A}_e$. The* force *of $<H, h>$ is $Force(A)$ = $min\{a_i$ such that $(\varphi_i, a_i) \in H\}$.*

Example 9 *In example 6, the force of the explanatory argument $<\{conference, conference \rightarrow Sydney\}, Sydney>$ is 0.8. Whereas, the force of the argument $<\{cancelled, cancelled \rightarrow \neg conference\}, \neg conference>$ is 0.4.*

Concerning the threats, things are different since a threat involves goals and beliefs. Intuitively, a threat is strong if, according to the most certain beliefs, it invalidates an important goal. A threat is weak if, according to the less certain beliefs, it invalidates a less important goal. In other terms,

the force of a threat represents to what extent the agent (the agent sending it or receiving it) is certain that it will violate its most important goals. Formally:

Definition 5 (Force of a threat) *Let $A = <H, h, \phi> \in \mathcal{A}_t$. The* force *of a threat A is $Force(A) = min(\alpha, \beta)$ such that $\alpha = min\{a_i$ such that $(\varphi_i, a_i) \in H$ and $(\phi, \beta) \in \mathcal{GO} \cup \mathcal{G}$.*

Note that when a threat is evaluated by the proponent (the agent presenting the threat), then $(\phi, \alpha) \in \mathcal{GO}$. However, when it is evaluated by its receiver, $(\phi, \alpha) \in \mathcal{G}$.

Example 10 *In example 3 the force of the threat $<\{\neg\ finish - work \rightarrow overtime\}, finish - work, \neg overtime>$ is min(1, 0.7) = 0.7.*

As for threats, rewards involve beliefs and goals. Thus, a reward is strong when it is for sure that it will contribute to the achievement of an important goal. It is weak if it is not sure that it will contribute to the achievement of a less important goal.

Definition 6 (Force of a reward) *Let $A = <H, h, \phi> \in \mathcal{A}_r$. The* force *of a reward A is $Force(A) = min(\alpha, \beta)$ such that $\alpha = min\{a_i$ such that $(\varphi_i, a_i) \in H$ and $(\phi, \beta) \in \mathcal{GO} \cup \mathcal{G}$.*

Example 11 *In example 5, the force of the reward $<\{finish - work \rightarrow high - budget, high - budget \rightarrow high - salary\}, finish - work, high - salary>$ is 0.6.*

The forces of the arguments makes it possible to compare different arguments as follows:

Definition 7 (Preference relation) *Let A_1 and A_2 be two arguments of $\mathcal{A}_x$. A_1 is* preferred *to A_2, denoted by $A_1 \succ A_2$, iff $Force(A_1) \geq Force(A_2)$.*

In fact, the forces of arguments will play two roles: in one hand they allow an agent to compare different threats or different rewards in order to select the "best" one. In the other hand, the forces are useful for determining the acceptable arguments among the conflicting ones.

Conflicts between arguments

Due to inconsistency in knowledge bases, arguments may be conflicting. In this section, we will show the different kinds of conflicts which may exist between arguments of the same nature and also between arguments of different natures.

Conflicts between explanatory arguments

In classical argumentation frameworks, different conflict relations between what we call in this paper explanatory arguments have been defined. The most common ones are the relations of *rebut* where two explanatory arguments support contradictory conclusions and the relation of *undercut* where the conclusion of an explanatory argument contradicts an element of the support of another explanatory argument.

Definition 8 *Let $<H, h>$, $<H', h'> \in \mathcal{A}_e$. $<H, h>$ $defeats_e$ $<H', h'>$ iff $\exists$ h" $\in H'$ such that $h \equiv \neg h$", or $h \equiv \neg h'$.*

Example 12 (Continued) *In example 6, the explanatory argument $<\{cancelled, cancelled \rightarrow \neg conference\}, \neg conference>$ undercuts the argument $<\{conference, conference \rightarrow Sydney\}, Sydney>$ whereas it rebuts the argument $<\{conference\}, conference>$.*

Conflicts between threats / rewards

Two arguments of type "threats" may be conflicting for one of the three following reasons:

- the support of an argument inferes the negation of the conclusion of the other argument. This case occurs when, for example, an agent P threatens C to do β if C refuses to do α, and at his turn, C threatens P to do δ if P does β.
- the threats support contradictory conclusions.
- the threatened goals are contradictory. Since a rational agent should have consistent goals, this case arises when the two threats are given by different agents.

As for threats, rewards may also be conflicting for one of the three following reasons:

- the support of an argument inferes the negation of the conclusion of the other argument. This occurs when an agent P promises to C to do β if C refuses to do α. C, at his turn, promises to P to do δ if P doesn't pursue β.
- the rewards support contradictory conclusions. This kind of conflict has no sense if the two rewards are constructed by the same agent. Because this means that the agent will contribute to the achievement of a goal of the other agent regardless what the value of h is. However, when the two rewards are given by different agents, this means that one of them wants h and the other $\neg h$ and each of them tries to persuade the other to change its mind by offering a reward.
- the rewarded goals are contradictory.

Formally:

Definition 9 *Let $<H, h, \phi>$, $<H', h', \phi'> \in \mathcal{A}_t$ (resp. $\in \mathcal{A}_r$). $<H', h', \phi'>$ $defeats_t$ $<H, h, \phi>$ (resp. $<H', h', \phi'>$ $defeats_r$ $<H, h, \phi>$) iff: $H' \vdash \neg h$, or $h \equiv \neg h'$, or $\phi \equiv \neg \phi'$.*

Note that the conflict relation between threats (or rewards) is generally symmetric.

Mixed conflicts

It is obvious that explanatory arguments can defeat threats and rewards. In fact, one can easily undercut an element used in the support of a threat or a reward. The defeat relation used in this case is the relation "undercut" defined above. An explanatory argument can also defeat a threat or a reward when the two arguments have contradictory conclusions. Finally, an explanatory argument may conclude the negation of the goal threatened (resp. rewarded) by the threat (resp. the reward). Formally:

Definition 10 *Let $<H, h> \in \mathcal{A}_e$ and $<H', h', \phi> \in \mathcal{A}_t$ (resp. $\in \mathcal{A}_r$). $<H, h>$ $defeats_m$ $<H', h', \phi>$ iff: $\exists h" \in H'$ such that $h \equiv \neg h"$ or $h \equiv \neg h'$ or $h \equiv \neg \phi$.*

Evaluation of Arguments

In classical argumentation, a basic argumentation framework is defined as a pair consisting of a set of arguments and a binary relation representing the defeasibility relationship between arguments. In such a framework, arguments are all considered as explanatory. However, in this paper we have argued that arguments may be of different natures. So the basic framework introduced initially by Dung in (Dung 1995) will be extended.

Definition 11 (Argumentation framework) *An* argumentation framework *is a tuple* $\langle \mathcal{A}_e, \mathcal{A}_t, \mathcal{A}_r, defeat_e, defeat_t, defeat_r, defeat_m \rangle$.

This framework will return three categories of arguments:

- The class of *acceptable* arguments. Indeed, the conclusions of acceptable explanatory arguments will hold and inferred from the bases. Conclusions of acceptable threats should also be considered. In fact, such threats are seen as serious ones. Finally, conclusions of acceptable rewards should be retained since the reward will be pursued.
- The class of *rejected* arguments. An argument is rejected if it is defeated by an acceptable one. Conclusions of rejected explanatory arguments will not be inferred from the bases. Rejected threats will not be considered since they are weak or not credible. Similarly, rejected rewards will be discarded since they are considered as weak.
- The class of arguments *in abeyance*. Such arguments are neither acceptable nor rejected.

In what follows, we will try to define what is an acceptable argument. Intuitively, it is clear that an argument which is not defeated at all will be accepted. $\mathcal{C}$ will denote the set of all the arguments of ($\mathcal{A}_e$, $\mathcal{A}_t$, $\mathcal{A}_r$) which are not defeated.
Due to the forces of the arguments, as in (Amgoud & Cayrol 2002a), we can accept some defeated arguments if they are stronger than any defeaters.

Definition 12 *The set of acceptable arguments is* $\mathcal{C}_{\succ} = \{A \in \mathcal{A}_x$ *such that* $\forall B \in \mathcal{A}_y, A \succ B\}$.

Illustrative example

Let us illustrate the proposed framework in a negotiation dialogue between a boss B, and a worker W. However, for the sake of simplicity, the strategy about decision moves is not discussed in detail. The knowledge base $\mathcal{K}_B$ of B is made of the following pieces of information, whose meaning is easy to guess ('overtime' is short for 'ask for overtime'): $\mathcal{K}_B$ = {(person-sick, 1), (person-sick $\rightarrow$ late-work, a_1), (late-work $\wedge$ $\neg$ overtime $\rightarrow$ $\neg$ finished-in-time, a_2), (overtime $\rightarrow$ finished-in-time, 1), ($\neg$ finished-in-time $\rightarrow$ penalty, 1), (overtime $\rightarrow$ pay $\vee$ free-day, 1), (pay $\rightarrow$ extra-cost, 1)} with $a_1 > a_2$.
Possible actions for B are represented by their effects under the form of fully certain propositions: $\mathcal{A}_B$ = {(T, 1), (overtime, 1), (pay, 1), (free-day, 1)}, where T denotes the tautology and corresponds to the result of the action 'do nothing'.
Goals of B are given by $\mathcal{G}_B$ = {($\neg$ penalty, b_1), ($\neg$ extra-cost, b_2), $\neg$ free-day, b_3)}, with $b_1 > b_2 > b_3$.
What he thinks are the goals of W are $\mathcal{GO}_B$ = {(pay, 1), ($\neg$ overtime, c)}.
On his side, W has the following bases: $\mathcal{K}_W$ = {(person-sick $\rightarrow$ late-work, d_1), (overtime $\rightarrow$ late-work, 1), (late-work $\wedge$ pay $\rightarrow$ overtime, d_1), (free-day $\rightarrow$ get-free-time, 1), (pay $\rightarrow$ get-money, 1), ($\neg$ late-work, d_2)}, with $d_1 > d_2$. $\mathcal{G}_W$ = {($\neg$ overtime $\vee$ pay, 1), (get-money, e_1), ($\neg$ overtime, e_2), (get-free-time, e_3)} with $e_1 > e_2 > e_3$, $\mathcal{GO}_W$ = {($\neg$ pay, f)}.
For the sake of simplicity, the set of possible actions of W is not used in the example.

Here it's a sketch of what can take place between B and W. In the current situation *(person-sick, 1)*, B is led to choose the actions 'overtime' and "free-day" (according to a regulation he knows in $\mathcal{K}_B$). Indeed it can be checked that this decision maximizes in $\mathcal{A}_B$ a pessimistic qualitative utility (Dubois *et al.* 1998); see (Dubois, Prade, & Sabbadin 2001) for axiomatic justifications. More precisely, "overtime" maximizes a such that ($\mathcal{K}_{Ba}$), *overtime* $\vdash (\mathcal{G}_B)_{\overline{1-a}}$, where ($\mathcal{K}_{Ba}$) is the set of formulas having a level of certainty at least equal to a, $(\mathcal{G}_B)_{\overline{1-a}}$ is the set of goals with a priority strictly greater than $1-a$.
Here $(\mathcal{K}_B)_{a_2}$, $overtime \vdash \neg penalty$ with $(\mathcal{G}_B)_{b_1}$ = {$\neg$ penalty}. If B does nothing (action (T,1)), $\mathcal{K}_B \vdash_{PL} (penalty, min(a_1, a_2))$ (where $\vdash_{PL}$ denotes the possibilistic logic consequence relation (Dubois, Lang, & Prade 1993). This would contradict his most prioritary goal in $\mathcal{G}_B$. The chosen action only contradicts his less prioritary goal, namely "free-day". B knows also that $overtime$ is a threat for W, but not so strong (c ¡ 1) according to $\mathcal{GO}_B$.
When W receives the command $overtime$, it challenges it since it believes $\neg overtime$ (indeed $\mathcal{K}_W \vdash_{PL} (\neg overtime, d_2)$), due to the argument {(overtime $\rightarrow$ late-work, 1), ($\neg$ late-work, d_2)}.
Then B provides the explanatory counter-argument {person-sick, person-sick $\rightarrow$ late-work}.
Then W accepts to revise his knowledge base by accepting (late-work, 1), since he ignored (person-sick, 1). Although "free-day" is a reward for him with strength e_3 (according to $\mathcal{K}_W$ and $\mathcal{G}_W$), he still does not endorse "overtime", which is thus not perceived as a threat for him. Indeed according to $\mathcal{K}_W$, the only case when he is obliged to accept "overtime" is under the two conditions "late-work" and "pay".
When B sees that W does not endorse "overtime", he regretfully proposes "pay" (since it violates his secondary goal), and considers that it is a strong "reward" for W (according to $\mathcal{GO}_B$). W feels B's offer a bit as a threat, that he cannot escape here by doing something), since it violates his third goal; it's also a reward since it pleases his three other goals!

Conclusion

Argumentation-based negotiation focuses on the necessity of exchanging arguments during a negotiation process. In fact, an offer supported by an argument has a better chance to be accepted by the other agent. In (Kraus, Sycara, & Evenchik 1998), a list of the different kinds of arguments that may be exchanged during a negotiation has been addressed. Among those arguments, there are the threats and the rewards. The authors have then tried to define how those arguments are generated. They presented that in terms of speech acts having pre-conditions. Later on in (Ramchurn,

Jennings, & Sierra 2003), the authors have tried to give a way for evaluating the force of threats and rewards. However no formalization of the different arguments has been given.

The aim of this paper is twice. First, it presents a logical framework in which the arguments are defined, the different conflicts which may exist between them are described, the force of each kind of arguments is defined in a clear way on the basis of the different bases of an agent and finaly the acceptability of the arguments is studied. This work can be seen as a first formalization of different kinds of arguments. This is beneficial both for negotiation dialogue and also for argumentation theory since in classical argumentation the nature of arguments is not taken into account or the arguments are supposed to have the same nature.

An extension of this work will be to study more deeply the notion of acceptability of such arguments. In this paper we have presented only the individual acceptability where only the direct defeaters are taken into account. However, we would like to investigate the notion of joint acceptability as defined by Dung in classical argumentation. We are also planning to investigate more deeply the language used in our framework. In fact, in this paper we have used a propositional language and thus no distinction is done between a fact and an action. Another perspective of this work is to investigate the integration of this framework in the more general architecture of a negotiation dialogue introduced in (Amgoud & Prade 2004).

References

Amgoud, L., and Cayrol, C. 2002a. Inferring from inconsistency in preference-based argumentation frameworks. *International Journal of Automated Reasoning* Volume 29, N2:125–169.

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

Amgoud, L., and Prade, H. 2004. Reaching agreement through argumentation: A possibilistic approach. In *9 th International Conference on the Principles of Knowledge Representation and Reasoning, KR'2004*.

Dubois, D.; Berre, D. L.; Prade, H.; and Sabbadin, R. 1998. Logical representation and computation of optimal decisions in a qualitative setting. In *15th National Conference on Artifical Intelligence (AAAI-98)*, 588–593.

Dubois, D.; Lang, J.; and Prade, H. 1991. A brief overview of possibilistic logic. In *Proc. of Symb. and Quanti. Approaches to Uncert., ECSQARU'91. LNCS 548.*, 53–57.

Dubois, D.; Lang, J.; and Prade, H. 1993. Possibilistic logic. *Handbook of Logic in Artificial Intelligence and Logic Programming* 3:439–513.

Dubois, D.; Prade, H.; and Sabbadin, R. 2001. *Decision-theoretic foundations of qualitative possibility theory*, volume 128. European Journal of Operational Research.

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.; Sycara, K.; and Evenchik, A. 1998. *Reaching agreements through argumentation: a logical model and implementation*, volume 104. Journal of Artificial Intelligence.

Parsons, S.; Sierra, C.; and Jennings, N. R. 1998. Agents that reason and negotiate by arguing. *Journal of Logic and Computation* 8(3):261—292.

Prakken, H., and Sartor, G. 1997. Argument-based extended logic programming with defeasible priorties. *Journal of Applied Non-Classical Logics* 7:25–75.

Rahwan, I.; Ramchurn, S. D.; Jennings, N. R.; McBurney, P.; Parsons, S.; and Sonenberg, L. 2004. Argumentation-based negotiation.

Ramchurn, S. D.; Jennings, N.; and Sierra, C. 2003. Persuasive negotiation for autonomous agents: a rhetorical approach. In *IJCAI Workshop on Computational Models of Natural Arguments*.