Handling threats, rewards and explanatory arguments in a unified setting

Leila Amgoud Henri Prade

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

Abstract. Current logic-based handling of arguments has mainly focused on explanation or justification-oriented purposes in presence of inconsistency. So only one type of argument has been considered, and 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* and *appeals*.
The purpose of this paper is to provide a logical setting 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 that negotiation dialogues involve not only agents' beliefs (of various strengths), but also their goals (having maybe different priorities), as well as the beliefs on the goals of other agents. In other words, from the different beliefs and goals bases maintained by agents, all the possible threats, rewards, explanations, appeals which are associated with them can be generated. It may also happen that an intended threat, or reward, is not perceived as such by the addressee and thus misses its target because the addresser misrepresents addressee's goals. The proposed approach accounts for that phenomenon. Finally, we show how to evaluate conflicting arguments of different types.

Key words: Argumentation, Negotiation.

1 INTRODUCTION

Argumentation is a promising approach for reasoning with inconsistent knowledge, based on the construction and the comparison of arguments. It may also be considered as a different method for handling uncertainty [9, 19, 24]. A basic idea behind argumentation is that it should be possible to say more about the certainty of a particular fact than just 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 have been developed in defeasible reasoning [1, 2, 13, 23, 25, 30] for generating and evaluating arguments. In that explanation-oriented perspective, only one type of argument has been considered. Namely, what is called here *explanatory* arguments.

Recent works on negotiation [4–6, 18, 22, 27–29] have argued that argumentation can play 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. 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*, and *appeals*. For example, if an agent receives a threat, this agent may accept the offer even if it is not really acceptable for it (because otherwise really important goals would be threatened).

Besides, evaluating and detecting threats also become an active research area both in military applications and intrusion detection for computer security, where various approaches have been proposed (e.g. [8, 15, 16]), including fuzzy set-based ones. However, the context of the study here is different and the emphasis is on the capacity of building verbal threats (as well as rewards) from available knowledge.

The purpose of this paper[1] 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, any possible threats, rewards, explanations, appeals, which are associated with them can be generated. Note that our weighting systems for threats and rewards reflect the certainty that they can take place and the importance of their consequences. However, they don't account for the propensity of the agent to act or not as it promises to do.

This paper is organized as follows: Section 2 discusses the different types of arguments identified in [18]. Section 3 introduces the logical language which will be used throughout the paper. Sections 4, 5 and 6 study the explanatory arguments (resp. threats and rewards). They present the formal definitions of each type of arguments as well as their strengths. In section 7 we discuss the different kinds of conflicts which may arise between the three types of arguments, and in section 8, we evaluate the acceptability of the arguments. An illustrative exam-

[1] This paper is an extended version of [7].

ple involving decision is provided in section 9. We compare our proposal with existing works in section 10. Finally, section 11 is devoted to some concluding remarks and perspectives.

2 TYPES OF ARGUMENTS

Arguments provide reasons for believing, justifications for acts, explanations of state of facts. In [31], it has been pointed out that it is not possible to present an exhaustive classification of arguments, since arguments must be interpreted and are effective within a particular context and domain. For example, when inferring from inconsistent knowledge bases, arguments aim at finding the most supported beliefs. However, during a negotiation the exchange of arguments may lead the agent which receives them to change its goals or preferences. In [17, 18, 21, 26], a list of the types of arguments, which are commonly thought to have persuasive force in human negotiations, has been identified and discussed. Six types of arguments are thus distinguished:

- Threats.
- Rewards.
- Appeal to past promise.
- Appeal to prevailing practice.
- Appeal to self-interest.
- Counter-examples.

In what follows we focus on the above list of arguments and we argue that three categories of arguments can be distinguished according to their logical definitions: *threats*, *rewards* and *explanatory arguments*. Indeed, a formal definition of threats and rewards requires two distinct bases: a knowledge base and a goals base. Whereas the definition of an explanatory argument requires only a knowledge base. Moreover, as emphasized in this paper, the definitions of threats and rewards have an *abductive* flavor, while the definition of explanatory arguments, which encompasses the four other kinds of arguments, is *deductive* in nature.

3 THE LOGICAL LANGUAGE

In what follows, $\mathcal{L}$ denotes a propositional language, $\vdash$ classical inference, and $\equiv$ logical equivalence. We suppose that we have two negotiating agents: P (called a proponent) and O (called an opponent). In all what follows, we suppose that P presents an argument to O. In a dialogue each agent plays these two roles in turn.

Each negotiating agent has got 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 [6].

$\mathcal{K}$ 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, each base is supposed to be equipped with a total preordering $\geq$.

$$a \geq b \text{ iff } a \text{ is at least as certain (resp. as preferred) as } b.$$

For encoding it, we use the set of integers $\{0, 1, \ldots, n\}$ as a linearly ordered scale, where n stands for the highest level of certainty or importance and '0' corresponds to the complete lack of certainty or importance. This means that the base $\mathcal{K}$ is partitioned and stratified into $\mathcal{K}_1$, ..., $\mathcal{K}_n$ ($\mathcal{K} = \mathcal{K}_1 \cup \ldots \cup \mathcal{K}_n$) such that all beliefs in $\mathcal{K}_i$ have the same certainty level and are more certain than beliefs in $\mathcal{K}_j$ where j < i. Moreover, $\mathcal{K}_0$ is not considered since it gathers formulas which are completely uncertain, and which are not at all beliefs of the agent.
Similarly, $\mathcal{GO} = \mathcal{GO}_1 \cup \ldots \cup \mathcal{GO}_n$ and $\mathcal{G} = \mathcal{G}_1 \cup \ldots \cup \mathcal{G}_n$ such that goals in $\mathcal{GO}_i$ (resp. in $\mathcal{G}_i$) have the same priority and are more important than goals in $\mathcal{GO}_j$ (resp. in $\mathcal{G}_j$ where j < i).

Note that some $\mathcal{K}_i$'s (resp. $\mathcal{G}_i$, $\mathcal{GO}_i$) may be empty if there is no piece of knowledge (resp. goal) corresponding to this level of certainty (resp. importance). For the sake of simplicity, in all our examples, we only specify the strata which are not empty. Both beliefs and goals are represented by propositional formulas of the language $\mathcal{L}$. Thus a goal is viewed as a piece of information describing a set of desirable states (corresponding to the models of the associated proposition) one of which should be reached. Let us consider the example of an agent who wants to buy a red car. In this case, the goal of this agent will be represented by $red \wedge car$.

4 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 a fact.

4.1 Logical definition

Such arguments have a *deductive* form. Indeed, from premises, a fact or a goal is entailed. Formally:

Definition 1 (Explanatory argument). *An* explanatory *argument is a pair <H, h> such that:*

1. $H \subseteq \mathcal{K}$,
2. $H \vdash h$,
3. *H is consistent and minimal (for set inclusion) among the sets satisfying the above conditions.*

$\mathcal{A}_e$ will denote the set of all the explanatory arguments that can be constructed from $\mathcal{K}$. H is the support *of the argument and h its* conclusion.

Note that h may be any propositional formula of the language $\mathcal{L}$. Let's illustrate the above definition on the following example.

Example 1 *Let us consider the case of an agent who wants to go to a conference in Sydney because he believes that it should obtain some money support for attending. $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_{n-1}$ such that $\mathcal{K}_n = \{Support \rightarrow GoSydney\}$, $\mathcal{K}_{n-1} = \{Support\}$.*

The agent wants to go to Sydney and justifies his goal by the following explanatory argument: <{Support, Support → GoSydney}, GoSydney>. Indeed, from the beliefs one can deduce GoSydney.

Note that the support of an explanatory argument is a subset of the beliefs of the agent (its base $\mathcal{K}$). In fact, the bases of goals are not considered when constructing arguments. The reason is that one should avoid *wishful thinking* as in the following example:

Example 2 *Let us again consider the case of an agent who wants to go to a conference. He believes that if there is enough money then he can go to the conference otherwise it will not be possible. These beliefs are encoded as follows: $\mathcal{K}_n = \{Money \rightarrow Conference\}$, $\mathcal{K}_{n-1} = \{\neg Money \rightarrow \neg Conference\}$ and $\mathcal{G}_n = \{Conference\}$.*
Using $\mathcal{K} \cup \mathcal{G}$, one can produce the following argument: <{Conference, ¬Money → ¬Conference}, Money>, which amounts to prove that one has money. However, this information cannot be deduced from $\mathcal{K}$ and may be completely wrong, since one takes for granted that the agent will really go to the conference (which in fact is only a wish).

4.2 Appeals and counter-examples

In [18] other types of arguments called *appeals* are also considered. We argue that the different forms of appeals (except the appeal to self-interest) can be modeled as explanatory arguments. In what follows, we will show through examples how appeals can be defined in this way.

Appeals to prevailing practice
During a negotiation, this kind of arguments is presented when the proponent agent believes that the opponent refuses to perform a requested action since it contradicts one of its own goals. The proponent provides then an example taken 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 should have performed the action successfully. Let's take the following example:

Example 3 *An agent P asks another agent O to make overtime. O refuses because he is afraid that this is prosecuted by law.*

P: You should make overtime.
O: No, this is prosecuted by law.
P: My colleague makes overtime and he never has problems with the law.

The different bases of O are: $\mathcal{K}_n^O = \{Overtime, Overtime \rightarrow ToBeProsecuted\}$, $\mathcal{G}_n^O = \{\neg ToBeProsecuted\}$ *and* $\mathcal{GO}_n^O = \emptyset$.
When the opponent O receives the offer 'overtime', he constructs an explanatory argument in favor of 'ToBeProsecuted': $<\{Overtime, Overtime \rightarrow ToBeProsecuted\}, ToBeProsecuted>$. *This argument confirms to him that his goal will be violated and he refuses the offer.*
The proponent P reassures him by telling that another colleague makes overtime and he never has problems with the law. The bases of P contains at least the following information: $\mathcal{K}_n^P = \{Overtime \wedge \neg ToBeProsecuted\}$, $\mathcal{G}^P = \{\emptyset\}$ *and* $\mathcal{GO}^P = \emptyset$.
In fact, he presents the following counter-argument: $<\{Overtime \wedge \neg ToBeProsecuted\}, \neg(Overtime \rightarrow ToBeProsecuted)>$. *This last argument is an appeal to prevailing practice.*

Appeals to past promise

It is very common that during a negotiation, an agent expects another agent to perform an action based on past promise. Let us illustrate it by the following example:

Example 4 *A child asks his mother to buy a gift for him and the mother refuses. Let's imagine the following dialogue:*

Child: I would like to have a gift since I succeeded at my examinations.
Mother: No.
Child: But you promised to buy something to me if I succeed at my examinations.

The bases of the child are: $\mathcal{K}_n = \{Success, Success \rightarrow Gift\}$, $\mathcal{G}_n = \{Gift\}$ *and* $\mathcal{GO} = \emptyset$. *The child's argument is then:* $<\{Success, Success \rightarrow Gift\}, Gift>$.

Counter-examples

A counter-example 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. This means that the agent P also maintains a belief base $\mathcal{KO}$ made of the propositions that represent beliefs of O according to P. $\mathcal{KO}$ is supposed to be layered in certainty levels as $\mathcal{K}$. Let's take an example:

Example 5 *Let's consider the following dialogue between an agent O who wants to buy a car and another agent P.*

O: I would like to buy a car of good quality. So I will not buy a second-hand one since such cars are of poor quality.

P: But, you have already bought a second hand car in a garage, which was fully satisfactory.

In this example, in order to try to convince O, P reminds it that it already bought a second hand car which was of good quality. The bases of P can be encoded as follows:
$\mathcal{K} = \emptyset$,
$\mathcal{KO}_n = \{SecondHand \rightarrow \neg Quality\}$,
$\mathcal{KO}_{n-1} = \{SecondHand \wedge Garage \wedge Quality\}$,
$\mathcal{G} = \emptyset$,
$\mathcal{GO}_n = \{Quality\}$.
The counter-example presented by P is $<\{SecondHand \wedge Garage \wedge Quality\}, \neg(SecondHand \rightarrow \neg Quality)>$.

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 more quickly lead the other agent to change its mind than an appeal to prevailing practice.

4.3 The strength of explanatory arguments

In [1], it has been argued that arguments may have forces of various strengths. These forces will play two roles: on the one hand they allow an agent to compare different arguments in order to select the 'best' ones. On the other hand, the forces are useful for determining the acceptable arguments among the conflicting ones.

Different definitions of the force of an argument have been proposed in [1]. Generally, the force of an argument can rely on the beliefs from which it is constructed. Indeed, explicit priorities between beliefs, or implicit priorities such as specificity, can be the basis for defining the force of an argument. However, different other aspects can be taken into account when defining the force of explanatory arguments. In particular, the length of the argument (in terms of the number of pieces of knowledge involved) may be considered since the shorter is the explanation, the better it is and the more difficult it is to challenge it (provided that it is based on propositions that are sufficiently certain).

When explicit priorities are given between the beliefs, such as certainty levels, the arguments using more certain beliefs are found stronger than arguments using less certain beliefs. The force of an explanatory argument corresponds to the *certainty level* of the less entrenched belief involved in the argument. In what follows, we consider the above definition of force. In the case of our stratified bases, the force of an argument corresponds to the smallest number of a stratum met by the support of that argument. Formally:

Definition 2 (Certainty level). *Let $\mathcal{K} = \mathcal{K}_1 \cup \ldots \cup \mathcal{K}_n$ be a stratified base, and $H \subseteq \mathcal{K}$. The certainty level of H, denoted Level(H) = $min\{j \mid 1 \leq j \leq n$ such that $H_j \neq \emptyset\}$, where H_j denotes $H \cap \mathcal{K}_j$.*

Note that $<H, h>$ is all the stronger as Level(H) has a large value.

Definition 3 (Force of an explanatory argument). *Let $A = <H, h> \in \mathcal{A}_e$. The* force *of A is Force(A) = Level(H).*

This definition agrees with the definition of an argument as a minimal set of beliefs supporting a conclusion. Indeed, when any member of this minimal set is seriously challenged, the whole argument collapses. This makes clear that the strength of the least entrenched argument fully mirrors the force of the argument whatever are the strengths of the other components in the minimal set.

Example 6 *In example 1, the force of the explanatory argument <{Support, Support → GoSydney}, GoSydney> is n-1.*

The forces of arguments make it possible to compare any pair of arguments. Indeed, arguments with a higher force are preferred. Formally:

Definition 4 (Preference relation between explanatory arguments). *Let $A, B \in \mathcal{A}_e$. A is preferred to B, denoted by $A \succ_e B$, iff Force(A) > Force(B).*

When two arguments go down to the same stratum, the above relation cannot compare them. In some situations, however, it is clear that one argument is better than the other. Let's assume a scale $\{0, 1, 2, 3, 4\}$ and let's suppose that the support of an argument A takes three formulas in the three following strata (4, 4, 1), and another argument B uses only two formulas in the strata (2, 1). According to the above definition, these two arguments have the same force, then they are not comparable. However, it is clear that the argument A is better than B since it uses more interesting formulas than B. To capture this idea of comparing the remaining formulas, a refinement has been proposed in [3]. The idea is to extend the concept of level to gradual levels as follows.

Definition 5 (Gradual certainty level). *Let $\mathcal{K} = \mathcal{K}_0 \cup \ldots \cup \mathcal{K}_n$ be a stratified base, and $H \subseteq \mathcal{K}$. For each $1 \leq k \leq n$, $Level_k(H) = Level(H_k \cup \ldots H_n) = min\ \{j \mid n \geq j \geq k$ and $H_j \neq \emptyset\}$ (with min $\emptyset$ taken equal to k).*

Property 1. Let $\mathcal{K} = \mathcal{K}_0 \cup \ldots \cup \mathcal{K}_n$ be a stratified base, and $H \subseteq \mathcal{K}$. Level(H) = $Level_1(H)$.

Using the above definition, arguments can be compared as follows:

Definition 6. *Let <H, h>, <H', h'> $\in \mathcal{A}_e$. <H, h> is preferred to <H', h'> iff $\exists$ $1 \leq k \leq n$ such that $Level_k(H) > Level_k(H')$ and for each $j < k$, $Level_j(H) = Level_j(H')$.*

Property 2. Let $A, B \in \mathcal{A}_e$. If A is preferred to B w.r.t the certainty level, then A is preferred to B w.r.t the gradual certainty level. The reverse does not always hold.

Note that the way definitions 5 - 6 refine definitions 2-4 does not account for the numbers of formulas inside the same stratum involved in an argument. For instance, no difference will be made by definitions 5 - 6 between an argument A with three formulas of respective levels (1,1,3) and an argument B with three formulas in levels (1,3,3), although the second may be found stronger. Namely, this suggests to use an idea that has been introduced in [14] for multiple criteria decision for refining minimum-based aggregation, and which is known as *leximin* [20].

Definition 7. *Let <H, h>, <H', h'> $\in \mathcal{A}_e$. Let (a_1, ..., a_r), (b_1, ..., b_s) be the vectors of certainty levels of the r (resp. s) formulas composing H (resp. H'), assumed to be increasingly ordered, i.e $a_1 \leq \ldots \leq a_r$ (resp. $b_1 \leq \ldots \leq b_s$). Assume that $r \leq s$.*
<H, h> is preferred to <H', h'> iff:

- *$a_1 > b_1$, or*
- *$\exists\ k \leq r$ such that $a_k > b_k$ and $\forall\ j < k$, $a_j = b_j$, or*
- *if $\forall\ 1 \leq j \leq r$, $a_j = b_j$, and (b_{r+1}, ..., b_s) $\neq$ (n, ..., n).*

Using the scale {0, 1, ... n}, this gives priority to the shortest arguments. Now the argument using formulas with the following certainty levels (1,3,3) is preferred to the one using (1,1,3). An argument using (1,3) is preferred to (1,3,3), and an argument using (1) is preferred to (1, 3) for $n = 4$. Note that definitions 5 - 6 find (1,3,3), (1,1,3) and (1,3) equivalent, while (1,3) is preferred to (1).

5 THREATS

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

- You should do 'a' otherwise I will do 'b',
- You should not do 'a' otherwise I will do 'b'.

The first case occurs when an agent P needs an agent O to do 'a' and O refuses. P, then threatens O to do 'b' which, according to its beliefs, will have bad consequences for O. Let us consider the following example.

Example 7 *Let's consider a mother who asks her child to carry out his school work.*

Mother: You should carry out your school work ('a').
Child: No, I don't want to.
Mother: You should carry out your school work otherwise I will not let you go to the party organized by your friend next week-end ('b').

The second kind of threats occurs when an agent O wants to do some action 'a', which is not acceptable for P. In this case, P threatens that if O insists to do 'a' then it will do 'b' which, according to P's beliefs, will have bad consequences for O. To illustrate this kind of threat, we consider the following example borrowed from [18].

Example 8
Labor union: We want a wage increase ('a').
Manager: I cannot afford that. If I grant this increase, I will have to lay off some employees ('b'). This will compensate for the higher operational cost that the increase will entail.

5.1 Logical definition

For a threat to be effective, it should be painful for its receiver and conflict with at least one of its goals. A threat is then made up of three parts: the *conclusion* that the agent who makes the threat wants and this can be seen as a goal of the proponent, the *threat* itself and finally the *threatened goal*. Moreover, it has an *abductive* form. Formally:

Definition 8 (Threat). *A* threat *is a triple* $<H, h, \phi>$ *such that:*

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

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

Note that in the above definition, h is a goal in $\mathcal{G}$, i.e. a goal of the agent which addresses the threat. In the case of a negotiation dialogue, for example, an agent P proposes an offer x and O refuses it. In this case P entices O in order to accept the offer otherwise it will do an action which may be painful for O. In this case h is $Accept(x)$, which is obviously a goal for P (see [6] for more details on agent communication languages in negotiation).

Moreover, such a definition allows h to be a proposition whose truth can be controlled by the threatened agent (e.g the result of an action), as well as a proposition which is out of the control of the agent. For instance, 'it rains and you are going to be wet'. We may however restrict the set where h is taken, in order to exclude the last case. Since we have imposed $h \in \mathcal{G}$, this forbids the situation where $h = \bot$ (the contradiction), which would correspond to the case of a gratuitous threat.

Note that the above definition captures the two forms of threats. Indeed, in

the first case (You should do 'a' otherwise I will do 'b'), h = 'a', and in the second case (You should not do 'a' otherwise I will do 'b'), $h = \neg a$. 'b' refers to an action which may be inferred from H. The formal definition of threats is then slightly more general.

Example 9 *As said in example 7, the mother threatens her child not to let him go to the party organized by his friend if he does't finish his school work. The mother is supposed to have the following bases:* $\mathcal{K}_m = \{\neg Work \rightarrow \neg Party\}$, $\mathcal{G}_m = \{Work\}$, $\mathcal{GO}_m = \{Party\}$.
The threat addressed by the mother to her child is formalized as follows: $<\{\neg Work \rightarrow \neg Party\}, Work, Party>$.

Let's now consider another dialogue between a boss and his employee.

Example 10
Boss: You should finish your work today.
Employee: No, I will finish it another day.
Boss: If you don't finish it you'll come this week-end to make overtime.
In this example, the boss has the three following bases: $\mathcal{K}_m = \{\neg\ FinishWork \rightarrow Overtime\}$, $\mathcal{G}_m = \{FinishWork\}$ *and* $\mathcal{GO}_m = \{\neg Overtime\}$.
The threat given by the boss is: $< \{\neg\ FinishWork \rightarrow Overtime\}, FinishWork, \neg Overtime>$.

5.2 The strength of threats

Compared to explanatory arguments, threats involve goals and beliefs. Thus, the force of a threat depends on two criteria: the *certainty level* of the beliefs used in that threat (i.e. the support), and the *importance* of the threatened goal. Moreover, when a threat is evaluated by the proponent (the agent presenting the threat), the threatened goal is in $\mathcal{GO}$. Formally:

Definition 9 (Force of a threat from the proponent's point of view). *Let* $A = <H, h, \phi> \in \mathcal{A}_t$. *The* force *of a threat* A *is a pair* $Force(A) = <\alpha, \beta>$ *such that:*

α = *Level(H).*
$\beta = j$ *such that* $\phi \in \mathcal{GO}_j$.

However, when a threat is evaluated by its receiver (opponent), the threatened goal is in $\mathcal{G}$. In fact, the threatened goal may or may not be a goal of the opponent.

Definition 10 (Force of a threat from the opponent's point of view). *Let* $A = <H, h, \phi> \in \mathcal{A}_t$. *The* force *of a threat* A *is a pair* $<\alpha, \beta>$ *such that:*

α = *Level(H).*
If $\phi \in \mathcal{G}_j$ *then* $\beta = j$, *otherwise* $\beta = 0$.

Intuitively, a threat is strong if, according to the most certain beliefs, it invalidates an important goal. A threat is weaker if it involves beliefs with a low certainty, or if it only invalidates a goal with low importance. In other terms, the force of a threat represents to what extent the agent sending it (resp. receiving it) is certain that it will violate the most important goals of the other agent (resp. its own important goals). This suggests the use of a *conjunctive* combination of the certainty of H and the priority of the most important threatened goal. Indeed, a fully certain threat against a very low priority goal is not a very serious threat.

Definition 11 (Conjunctive combination). *Let A, $B \in \mathcal{A}_t$ with $Force(A) = <\alpha, \beta>$ and $Force(B) = <\alpha', \beta'>$. A is* stronger *than B, denoted by $A \succ_t B$, iff $min(\alpha, \beta) > min(\alpha', \beta')$.*

Example 11 *Assume the following scale $\{0, 1, 2, 3, 4, 5\}$. Let us consider two threats A and B whose forces are respectively $(\alpha, \beta) = (3, 2)$ and $(\alpha', \beta') = (1, 5)$. In this case the threat A is* stronger *than B since $min(3, 2) = 2$, whereas $min(1, 5) = 1$.*

However, a simple conjunctive combination is open to discussion, since it gives an equal weight to the importance of the goal threatened and to the certainty of the set of beliefs that establishes that the threat takes place. Indeed, one may feel less threatened by a threat that is certain but has 'small' consequences, than by a threat which has a rather small plausibility, but which concerns a very important goal. This suggests to use a weighted minimum aggregation as follows:

Definition 12 (Weighted conjunctive combination). *Let A, $B \in \mathcal{A}_t$ with $Force(A) = <\alpha, \beta>$ and $Force(B) = <\alpha', \beta'>$. A is* stronger *than B, denoted by $A \succ_t B$, iff $min(max(\lambda, \alpha), \beta) > min(max(\lambda, \alpha'), \beta')$, where λ is the weight that discounts the certainty level component.*

The larger λ is, the smaller the role of α in the evaluation.

Property 3. The conjunctive combination is recovered when the value of λ is minimal.

Example 12 *Assume the following scale $\{0, 1, 2, 3, 4, 5\}$. Let us consider two threats A and B whose forces are respectively $(\alpha, \beta) = (5, 2)$ and $(\alpha', \beta') = (2, 5)$. Using a simple conjunctive combination, they both get the same evaluation 2. Taking $\lambda = 3$, we have $min(max(3, 5), 2) = 2$ and $min(max(3, 2), 5) = 3$. Thus B is stronger than A as expected.*

The above approach assumes the commensurateness of three scales, namely the certainty scale, the importance scale, and the weighting scale. This requirement is questionable in principle. If this hypothesis is not made, one can still define a relation between threats as follows:

Definition 13. *Let A, $B \in \mathcal{A}_t$ with $Force(A) = <\alpha, \beta>$ and $Force(B) = <\alpha', \beta'>$. A is* stronger *than B iff:*

– $\beta > \beta'$ *or,*
– $\beta = \beta'$ *and* $\alpha > \alpha'$*.*

This definition also gives priority to the importance of the threatened goal, but is less discriminating than the previous one.

6 REWARDS

During a negotiation an agent P can entice agent O in order that it does 'a' by offering to do an action 'b' as a reward. Of course, agent P believes that 'b' will contribute to the goals of O. 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 'a' then I will do 'b'.
– If you do not do 'a' then I will do 'b'.

Let's illustrate the notion of reward by the following example.

Example 13 *In this example, a seller proposes to offer a set of blank CDs to a customer if this last accepts to buy a computer.*

Seller: This computer is very powerful ('a').
Customer: No I don't want it.
Seller: If you buy it I will offer you a set of blank CDs ('b').

6.1 Logical definitions

Formally, a reward is defined as follows:

Definition 14 (Reward). *A* reward *is a triple* $<H, h, \phi>$ *such that:*

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

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

As for threats, h is an element in $\mathcal{G}$, i.e a goal of the agent. Note that the above definition captures the two forms of rewards. Indeed, in the first case (If you do 'a' then I will do 'b'), $h =$ 'a', and in the second case (If you do not do 'a' then I will do 'b'), $h = \neg a$.

Example 14 *Let's consider the example of a boss who promises one of his employee to increase his salary.*

Boss: You should finish this work ('a').
Employee: No I can't.
Boss: If you finish the work I promise to increase your salary ('b').

The boss has the following bases: $\mathcal{K}_n = \{FinishWork \rightarrow IncreasedBenefit\}$, $\mathcal{K}_{n-1} = \{IncreasedBenefit \rightarrow HigherSalary\}$, $\mathcal{G}_n = \{FinishWork\}$ *and* $\mathcal{GO}_n = \{HigherSalary\}$.
The boss presents the following reward in favor of its request 'FinishWork': $<\{FinishWork \rightarrow HighBenefit, HighBenefit \rightarrow HighSalary\}, FinishWork, HighSalary>$.

Threats are sometimes thought as negative rewards. This is reflected by the parallel between the two definitions which basically differ in the third condition.

Property 4. Let $<\mathcal{K}, \mathcal{G}, \mathcal{GO}>$ be three bases of agent P. If $h \in \mathcal{G} \cup \mathcal{GO}$, then $<\emptyset, h, h>$ is both a reward and a threat.

The above property says that if h is a *common goal* of the two agents P and O, then $<\emptyset, h, h>$ can be both a reward and a threat, since the common goals jointly succeed or fail. This is either both a reward and a self-reward, or a threat or a self-threat for P.

In [18], 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 O's goals. In fact, this case may be seen as a *self-reward* and consequently it is a particular case of rewards.

Let us emphasize that a threat or a reward cannot be reduced to an explanatory argument as can be already seen on the definitions. On the one hand, explanatory arguments may lead the other agent to revise its beliefs / goals (they affect the mental states of the agent), while threats or rewards may encourage or refrain the agent to do something. On the other hand, the key entailment condition in the definition of threat, reward and explanatory arguments allows the following respective readings, H threatens ϕ provided $\neg h$, H rewards ϕ provided h and finally H explains h. Despite this apparent formal similarity, the two first expressions should be understood in a reverse way from an explanatory perspective. Indeed, in case of a threat or a reward this is rather the pair (H, ϕ) (although ϕ is the consequence of the entailment) which provides a kind of abductive explanation for h. Moreover, another important feature of definitions 8 and 14 is the requirement that ϕ belongs to $\mathcal{GO}$ which is distinct from $\mathcal{K}$ from which H is taken.

6.2 The strength of rewards

As for threats, rewards involve beliefs and goals. Thus, the force of a reward depends also on two criteria: the certainty level of its support and the importance of the rewarded goal. Moreover, when a reward is evaluated by the proponent (the agent presenting the reward), the rewarded goal is in $\mathcal{GO}$.

Definition 15 (Force of a reward from the proponent's point of view). *Let $A = <H, h, \phi> \in \mathcal{A}_r$. The* force *of a reward A is a pair $Force(A) = <\alpha, \beta>$ such that:*

α = *Level(H).*
$\beta = j$ *such that* $\phi \in \mathcal{GO}_j$.

However, when a reward is evaluated by its receiver (opponent), the rewarded goal is in $\mathcal{G}$. In fact, if the proponent does not misrepresent the goals of the opponent, the rewarded goal should be a goal of the opponent.

Definition 16 (Force of a reward from the opponent's point of view). *Let $A = <H, h, \phi> \in \mathcal{A}_r$. The* force *of a reward A is a pair $<\alpha, \beta>$ such that:*

α = *Level(H).*
If $\phi \in \mathcal{G}_j$ *then* $\beta = j$, *otherwise* $\beta = 0$.

Example 15 *In example 14, the force of the reward* $<\{FinishWork \rightarrow HighBenefit, HighBenefit \rightarrow HighSalary\}, FinishWork, HighSalary>$ *is* $<n-1, n>$.

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 an important goal, or if it is certain that it will only enable the achievement of a non very important goal. Formally:

Definition 17 (Conjunctive combination). *Let A, B be two rewards in $\mathcal{A}_r$ with $Force(A) = <\alpha, \beta>$ and $Force(B) = <\alpha', \beta'>$.*
A is preferred *to B, denoted by $A \succ_r B$, iff $min(\alpha, \beta) > min(\alpha', \beta')$.*

However, as for threats, a simple 'min' combination is open to discussion, since it gives an equal weight to the importance of the rewarded goal and to the certainty of the set of beliefs that establishes that the reward takes place. Indeed, one may feel less rewarded by a reward that is certain but has 'small' consequences, than by a reward which has a rather small plausibility, but which concerns a very important goal. This suggests to use a weighted minimum aggregation as follows:

Definition 18 (Weighted conjunctive combination). *Let A, $B \in \mathcal{A}_r$ with $Force(A) = <\alpha, \beta>$ and $Force(B) = <\alpha', \beta'>$.*
A is stronger *than B, denoted by $A \succ_r B$, iff $min(max(\lambda, \alpha), \beta) > min(max(\lambda, \alpha'), \beta')$, where λ is the weight that discounts the certainty level component.*

The larger λ is, the smaller the role of α in the evaluation.

Property 5. The 'min' combination is recovered when the value of λ is minimal.

In some situations, an agent may prefer a reward which is sure, even if the rewarded goal is not very important for it, than a reward that is uncertain but has very 'valuable' consequences. This suggests to use weighted minimum aggregation giving priority to the certainty component of the force, as follows:

Definition 19. *Let A, $B \in \mathcal{A}_r$ with $Force(A) = <\alpha, \beta>$ and $Force(B) = <\alpha', \beta'>$.*
A is stronger *than B, denoted by $A \succ_t B$, iff $min(\alpha, max(\lambda, \beta)) > min(\alpha', max(\lambda, \beta'))$, where λ is the weight that discounts the importance of the goal.*

Finally, as for threats, if there is no commensurateness of the three scales, we can still be able to compare two rewards as follows:

Definition 20. *Let A, $B \in \mathcal{A}_r$ with $Force(A) = <\alpha, \beta>$ and $Force(B) = <\alpha', \beta'>$. A is* stronger *than B iff:*

- $\beta > \beta'$ *or,*
- $\beta = \beta'$ *and* $\alpha > \alpha'$.

This definition also gives priority to the importance of the rewarded goal. In the case of an agent which prefers certain rewards even if the rewarded goals are not very important, we can have the following preference relation.

Definition 21. *Let A, $B \in \mathcal{A}_r$ with $Force(A) = <\alpha, \beta>$ and $Force(B) = <\alpha', \beta'>$. A is* stronger *than B iff:*

- $\alpha > \alpha'$ *or,*
- $\alpha = \alpha'$ *and* $\beta > \beta'$.

7 CONFLICTS BETWEEN ARGUMENTS

Due to the presence of 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.

7.1 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 *rebuttal* 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. Formally:

Definition 22. *Let $<H, h>$, $<H', h'> \in \mathcal{A}_e$.*

- $<H, h>$ undercuts $<H', h'>$ *iff* $\exists\ h'' \in H'$ *such that* $h \equiv \neg h''$.
- $<H, h>$ rebuts $<H', h'>$ *iff* $h \equiv \neg h'$.

Example 16 *Let us consider a variant of example 1. We suppose that the agent P wants to go to Sydney because there is a conference there. However, he believes that the conference is canceled. The different bases of P are encoded as follows: $\mathcal{K} = \mathcal{K}_n \cup \mathcal{K}_{n-1}$ such that $\mathcal{K}_n = \{Canceled, Canceled \rightarrow \neg Conference\}$, $\mathcal{K}_{n-1}$*

$= \{Conference,\ Conference \rightarrow GoSydney\}$, $\mathcal{G}_n = \{GoSydney\}$ and $\mathcal{GO} = \emptyset$.
The explanatory argument $<\{Canceled, Canceled \rightarrow \neg Conference\}, \neg Conference>$ undercuts the argument $<\{Conference, Conference \rightarrow GoSydney\}, GoSydney>$ whereas it rebuts the argument $<\{Conference\}, Conference>$.

In [1], it has been shown that the "undercut" relation captures all the different conflicts which may exist between arguments. Moreover, it is sufficient for handling inconsistency in propositional knowledge bases. In what follows, we will only consider that "undercut" relation. We bring together the undercut relation and the preference relation $\succ_e$ between arguments in a unique relation as follows:

Definition 23. *Let $<H, h>$, $<H', h'> \in \mathcal{A}_e$.*
$<H, h>$ $defeats_e$ $<H', h'>$ iff:

1. *$<H, h>$ undercuts $<H', h'>$ and*
2. *not ($<H', h'> \succ_e <H, h>$)*

In other terms, this means that an argument is defeated if it is undercut and it is not stronger than its undercutting argument.

7.2 Conflicts between threats / rewards

Two arguments of 'threat' type may be conflicting for one of the three following reasons:

- the support of an argument infers the negation of the conclusion of the other argument. This case occurs when, for example, an agent P threatens O to do 'b' if O refuses to do 'a', and at his turn, O threatens P to do 'c' if P does 'b'.
- the threats support contradictory conclusions. This case occurs, for example, when two agents P and O have contradictory purposes.
- the threatened goals are contradictory. Since a rational agent should have consistent goals, the base $\mathcal{GO}$ should be as well consistent, and thus 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 infers the negation of the conclusion of the other argument. This occurs when an agent P promises to O to do 'b' if O refuses to do 'a'. C, at his turn, promises to P to do 'c' if P does not pursue 'b'.
- 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 by two agents are contradictory.

Formally:

Definition 24. *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* $\phi \equiv \neg \phi'$, *and*
– *not* ($<H, h, \phi> \succ_t <H', h', \phi'>$) *(resp. not* ($<H, h, \phi> \succ_r <H', h', \phi'>$))

7.3 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. Lastly, an explanatory argument may conclude the negation of the goal threatened (resp. rewarded) by the threat (resp. the reward). Formally:

Definition 25. *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*
– $h \equiv \neg \phi$.

Property 6. The conflict relation given in definition 25 is asymetric.

Note that the force of the arguments is not taken into account when defining the relation "$defeat_m$". The reason is that firstly, the two arguments are of different nature. The force of explanatory arguments involves only beliefs while the the force of threats (resp. rewards) involves beliefs and goals. Secondly, beliefs have priority over goals since it is beliefs which determine whether a given goal is justified and feasible.

8 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 defeasible 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 nature. So the basic framework introduced initially by Dung in [13] will be extended as follows.

Definition 26 (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 gives rise to three categories of arguments:

- The class of *acceptable* arguments. Indeed, the conclusions of acceptable explanatory arguments will be inferred from the bases. Conclusions of acceptable threats should also be taken into account. In fact, such threats are very *serious*. Finally, conclusions of acceptable rewards should be retained since the reward may 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 taken into account 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.

Let us define what is an acceptable argument. Intuitively, it is clear that an argument which is not defeated (w.r.t $defeats_x$ with $x \in \{t, r, e, m\}$) will be accepted. Formally:

Definition 27 (Acceptable explanatory arguments). *Let* $\langle \mathcal{A}_e, \mathcal{A}_t, \mathcal{A}_r, defeat_e, defeat_t, defeat_r, defeat_m \rangle$ *be an argumentation framework. The* set of acceptable *explanatory arguments is*

$$\mathcal{S}_e = \{A \in \mathcal{A}_e | \nexists B \in \mathcal{A}_e, B \ defeats_e \ A\}.$$

An argument $A \in \mathcal{A}_e$ *is* acceptable *iff* $A \in \mathcal{S}_e$.

Similarly, acceptable threats and rewards can be defined.

Definition 28 (Acceptable threats). *Let* $\langle \mathcal{A}_e, \mathcal{A}_t, \mathcal{A}_r, defeat_e, defeat_t, defeat_r, defeat_m \rangle$ *be an argumentation framework. The* set of acceptable *threats is*

$$\mathcal{S}_t = \{A \in \mathcal{A}_t | \nexists B \in \mathcal{A}_t(resp.\mathcal{A}_e), B \ defeats_t \ (resp. \ defeats_m) \ A\}.$$

A threat $A \in \mathcal{A}_t$ *is* acceptable *iff* $A \in \mathcal{S}_t$.

Acceptable threats are the ones which are not defeated by another threat or by an explanatory argument.

Definition 29 (Acceptable rewards). *Let* $\langle \mathcal{A}_e, \mathcal{A}_t, \mathcal{A}_r, defeat_e, defeat_t, defeat_r, defeat_m \rangle$ *be an argumentation framework. The* set of acceptable *rewards is*

$$\mathcal{S}_r = \{A \in \mathcal{A}_r | \nexists B \in \mathcal{A}_r(resp.\mathcal{A}_e), B \ defeats_r \ (resp. \ defeats_m) \ A\}.$$

A threat $A \in \mathcal{A}_r$ *is* acceptable *iff* $A \in \mathcal{S}_r$.

Acceptable rewards are the ones which are not defeated by another reward or by an explanatory argument.

9 ILLUSTRATIVE EXAMPLE

Let us illustrate the proposed framework in a negotiation dialogue between a boss B, and a worker W. Each of them maintains the three bases $\mathcal{K}$, $\mathcal{G}$, $\mathcal{GO}$ corresponding to his own knowledge, preferences and beliefs on the goals of the other, and is supposed to have a set of possible actions that he may perform.

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 the decision 'overtime' maximizes in $\mathcal{A}_B$ a pessimistic qualitative utility [10]; see [12] for axiomatic justifications. More precisely, 'overtime' maximizes a such that

$$(\mathcal{K}_{Ba}), \mathit{overtime} \vdash (\mathcal{G}_B)_{\overline{m(a)}}$$

where m is the order reversing map of the scale (if the scale is $\{0, 1, \ldots, n\}$ then $m(a) = n - a$), and $(\mathcal{K}_{Ba})$ is the set of formulas having a level of certainty at least equal to a, $(\mathcal{G}_B)_{\overline{m(a)}}$ is the set of goals with a priority strictly greater than $m(a)$. The idea behind this definition is the following: from a pessimistic

point of view, a good decision is such that, taking the most certain part of the available knowledge into account, it entails that any goals having high priority are satisfied.

Here $(\mathcal{K}_B)_{a_2}$, $overtime \vdash \neg penalty$ with $(\mathcal{G}_B)_{b_1} = \{\neg \text{ penalty}\}$. 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 [11]. This would contradict his most prioritary goal in $\mathcal{G}_B$. The chosen action 'overtime' 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!

10 RELATED WORKS

The idea of threat has been somewhat pervasive in the decision under risk literature for a long time, since the high plausibility of a bad output in relation with the choice of an act can be indeed perceived as a threat. This is particularly the case when the threat is caused by some agent who may be suspected to act intentionally. The idea of threat evaluation is especially present in two types of information engineering applications, intrusion detection in computer security, and military target analysis. Such an evaluation takes into account how certain the threat is and how important it is if it takes place. The use of fuzzy logic-based techniques has been proposed for both types of applications [8, 15, 16].

In this paper, we are more concerned by the expression of a threat as a *special type of argument* and how it is perceived by another agent, as well as by the dual notion of reward. For that purpose, We use a graded view of uncertainty

and importance. However the proposed approach substantially differs from the ones proposed in [18, 29]. In these works, threats, rewards and appeals are considered as *persuasive particles* that agents can use in a negotiation. They claim that these particles are not arguments, but *speech acts* having preconditions and post-conditions. The preconditions represent the conditions that should be satisfied before sending a given particle. They are expressed through beliefs of the proponent about the value of the proposal (the conclusion h in our model). The post-conditions represent the consequences of that particle. In fact, they result in new beliefs being added to the opponent's beliefs base.
For instance, for an agent P to threaten an agent O, three preconditions must be satisfied:

1. P must believe that O prefers staying in the current state than enacting the proposal (the h in our model). This corresponds to the fact that an agent P presents an offer which is refused by the opponent.
2. P believes O will prefer to stay in the current state to having the threat effected. In our model, this is captured by the fact that a goal of the opponent will be violated if the threat takes place.
3. P must believe that the state brought about by the threat is less preferred by O than the state brought about by the proposal.

A state is defined as a pair of the beliefs of all the negotiating agents and a fully observable environment state. This hypothesis is very strong since it is supposed that the current state is known to all the agents, and that when generating a threat, an agent takes into account the beliefs of all the agents.
Each agent is equipped with two valuation functions taking their values in the interval [0, 1]. The first one affects a value to each state (representing the desirability of the state), and the second one measures the expected value of an action (the conclusion h in our model) in a given state.
Concerning rewards, for an agent P to reward an agent O, three preconditions must be satisfied:

1. P must believe that enacting the proposal h is less preferred by O to staying in the current state. This means that the agent O has already refused the offer h made by the agent P.
2. P believes that O can be rewarded with a more preferred alternative to the proposal h. In our model, this corresponds to the satisfaction of a goal of O if the reward takes place.
3. P must believe that the state brought about by the reward is more preferred by O than the state brought about by the proposal.

In sum, in these works threats and rewards are considered as any other speech acts. It is not clear how their forces are evaluated, nor how they can be defeated. However, these works are useful once our argumentation framework is integrated in an architecture of negotiation dialogue.

11 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 [18], a list of the different kinds of arguments that may be exchanged during a negotiation has been addressed. Among those arguments, there are threats and 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 [29], 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 twofold. Firstly, it presents a logical framework in which the arguments are defined. Moreover, the different conflicts which may exist between these arguments are described. Different criteria for defining the force of each kind of arguments are also proposed. We show clearly that the force of an explanatory argument depends on the beliefs from which that argument is built, whereas the force of threats or rewards depends on the beliefs of the agent and on its goals. Since arguments may be conflicting we have studied their acceptability.
Secondly, the work presented here 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. Another extension consists of studying the properties of the argumentation framework for any of the preference relations presented here. 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, which creates some limitations in the expressiveness of the definitions. Another perspective of this work is to investigate the integration of this framework in the more general architecture of a negotiation dialogue introduced in [6].

References

1. L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *International Journal of Automated Reasoning*, Volume 29, N2:125–169, 2002.
2. L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, Volume 34:197–216, 2002.

3. L. Amgoud, C. Cayrol, and D. LeBerre. Comparing arguments using preference orderings for argument-based reasoning. In *Proceedings of the 8th International Conference on Tools with Artificial Intelligence*, pages 400–403, 1996.
4. L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings of the International Conference on Multi-Agent Systems*, pages 31–38, Boston, MA, 2000.
5. L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In *Proceedings of the 14th European Conference on Artificial Intelligence*, 2000.
6. L. Amgoud and H. Prade. Reaching agreement through argumentation: A possibilistic approach. In *9 th International Conference on the Principles of Knowledge Representation and Reasoning*, Whistler, Canada, 2004.
7. L. Amgoud and H. Prade. Formal handling of threats and rewards in a negotiation dialogue. In *Proceedings of the 4th International joint Conference on Autonomous Agents and Multi-Agent Systems*, 2005.
8. A. Berrached, M. Beheshti, A. de Korvin, and R. Al. Applying fuzzy relation equations to threat analysis. In *Proc. 35th Annual Hawaii International Conference on System Sciences, Volume 2*, pages 50–54, 2002.
9. P. R. Cohen. *Heuristic Reasoning about uncertainty: An Artificial Intelligence approach.* Pitman Advanced Publishing Program, 1985.
10. D. Dubois, D. Le Berre, H. Prade, and R. Sabbadin. Logical representation and computation of optimal decisions in a qualitative setting. In *15th National Conference on Artifical Intelligence (AAAI-98)*, pages 588–593, 1998.
11. D. Dubois, J. Lang, and H. Prade. Possibilistic logic. *Handbook of Logic in Artificial Intelligence and Logic Programming*, 3:439–513, 1993.
12. D. Dubois, H. Prade, and R. Sabbadin. *Decision-theoretic foundations of qualitative possibility theory*, volume 128. European Journal of Operational Research, 2001.
13. 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.
14. H. Fargier, J. Lang, and T. Schiex. Selecting preferred solutions in fuzzy constraint satisfaction problems. In *Proceedings of the 1st European Congress on Fuzzy and Intelligent Technologies, EUFIT'93*, pages 1128–1134, 1993.
15. E. Hamed, J. Graham, and A. Elmaghraby. Computer system threat evaluation. In *Proc. 10th International Conference on Intelligent Systems. Washington, DC, International Society for Computers and Their Applications, Raleigh, NC.*, pages 23–26, 2001.
16. E. Hamed, J. Graham, and A. Elmaghraby. Fuzzy threat evaluation in computer security. In *Proc. International Conference on Computers and Their Applications. San Francisco, CA: International Society for Computers and Their Applications, Raleigh, NC*, pages 389–393, 2002.
17. M. Karlins and H. I. Abelson. Persuasion: How opinions and attitudes are changed. *Spinger Publishing Company, Inc.*, 1970.
18. S. Kraus, K. Sycara, and A. Evenchik. *Reaching agreements through argumentation: a logical model and implementation*, volume 104. Journal of Artificial Intelligence, 1998.
19. P. Krause, S. Ambler, M. Elvang-Gøransson, and J. Fox. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence*, 11:113–131, 1995.
20. H. Moulin. Axioms of cooperative decision making. *Cambridge University Press, Cambridge, UK*, 1988.
21. D. J. O'Keefe. Persuasion: Theory and research. *SAGE Publications*, 1990.

22. S. Parsons, C. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261—292, 1998.
23. J. L. Pollock. How to reason defeasibly. *Journal of Artificial Intelligence*, 57:1–42, 1992.
24. J. L. Pollock. Defeasible reasoning with variable degrees of justification. *Journal of Artificial Intelligence*, 333:233–282, 2001.
25. H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorties. *Journal of Applied Non-Classical Logics*, 7:25–75, 1997.
26. D. G. Pruitt. Negotiation behavior. *Academic Press, New York*, 1981.
27. I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. Argumentation-based negotiation. *Knowledge engineering review*, 2004.
28. I. Rahwan, L. Sonenberg, and F. Dignum. Towards interest-based negotiation. In *AAMAS'2003*, 2003.
29. S. D. Ramchurn, N. Jennings, and C. Sierra. Persuasive negotiation for autonomous agents: a rhetorical approach. In *IJCAI Workshop on Computational Models of Natural Arguments*, 2003.
30. G.R. Simari and R.P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Journal of Artificial Intelligence*, 53:125–157, 1992.
31. S. Toulmin, R. Reike, and A. Janik. An introduction to reasoning. *Macmillan Publishing Company, Inc.*, 1979.