Electronic Notes in Theoretical Computer Science 70 No. 5 (2002)
URL: http://www.elsevier.nl/locate/entcs/volume70.html 16 pages

Argumentative Agent Deliberation, Roles and Context

Antonis C. Kakas [1]

Department of Computer Science
University of Cyprus
P.O.Box 20537, CY-1678 Nicosia, Cyprus

Pavlos Moraïtis [2]

Department of Computer Science
University of Cyprus
P.O.Box 20537, CY-1678 Nicosia, Cyprus

Abstract

This paper presents an argumentation based framework to support an agent's deliberation process for drawing conclusions under a given policy. The argumentative policy of the agent is able to take into account the roles agents can have within a context pertaining to an environment of interaction. The framework uses roles and context to define policy preferences at different levels of deliberation allowing a modular representation of the agent's knowledge that avoids the need for explicit qualification of the agent's decision rules. We also employ a simple form of abduction to deal with the incompleteness and evolving nature of the agent's knowledge of the external environment and illustrate how an agent's self deliberation can affect the mode of interaction between agents. The high degree of modularity of the framework gives it a simple computational model in which the agent's deliberation can be naturally implemented.

1 Introduction

Autonomous agents need to make decisions under complex preference policies that take into account different factors. In general, these policies have a dynamic nature and are influenced by the particular state of the environment in which the agent finds himself. The agent's decision process needs to be able

[1] Email: antonis@ucy.ac.cy
[2] Email: moraitis@ucy.ac.cy

to synthesize together, in a modular way, different aspects of his preference policy and to adapt to new input from the current environment.

In this paper we present an argumentation based framework to support an agent's self deliberation process for drawing conclusions under a given policy. We will consider and extend an argumentation framework developed over the last decade as a result of a series of studies [11,8,7,10,9,6] on the links of argumentation to non-monotonic reasoning. This framework, called Logic Programming without Negation as Failure ($LPwNF$), was proposed originally in [9] and can be seen as a realization of the more abstract frameworks of [7,4]. The abstract attacking relation, i.e. its notion of argument and counter-argument, is realized through monotonic proofs of contrary conclusions and a priority relation on the sentences of the theory that make up these proofs. We extend the framework, following the more recent approach of other works [19,5] to allow this priority relation and thus the attacking relation to be dynamic, making the framework more suitable for the application of agent self deliberation.

In this work, we will consider that an agent is composed of a set of modules each of them being responsible for a particular capability (e.g. problem solving, cooperation, communication, etc.), and all together implementing the agent's overall behavior. Within this the proposed argumentative deliberation model can be used in order to implement the various decision making processes needed by different modules of an agent. For example, the decision for the choice and achievement of a goal (within the problem solving module) or the decision for the choice of the appropriate partners according to a specific cooperation protocol (within the cooperation module), etc.

Our argumentation framework captures agent deliberation in a dynamic external environment. In particular, we will examine the argumentative deliberation of an agent according to a given decision policy on a domain of interest that takes into account the roles filled by the agents and the context of the external environment.

Over the last few years argumentation is becoming increasingly important in agent theory. Several works have proposed argumentation models in the multi-agent field [22,21,17,15,3,1,2]. Our work can be seen as bringing together work from [21,2] who have suggested that roles can affect an agent's argumentation, especially within the context of a dialogue, and work from [19,5] who have shown the need for dynamic priorities within an argumentation framework when we want to apply this to formalize law and other related problems. In this paper, we put together these ideas proposing a new argumentation framework for agent deliberation obtained by extending the argumentation framework of ($LPwNF$) [9,6] to include dynamic priorities. We also employ a simple form of abduction to deal with the incompleteness and evolving nature of the agent's knowledge of the external environment.

We show how our framework can encompass the influence that the different relative roles of interacting agents and the context of the particular interaction can have on the deliberation process of the agents. Roles and context define in a natural way dynamic priorities on the argumentative decision rules of the agent at two different levels in the deliberation process. These priorities are represented within the overall argumentation theory of the agent in two corresponding modular parts.

Section 2 reviews the basic argumentation framework of $LPwNF$. Section 3, extends this framework to allow dynamic priorities and formulates the general framework of argumentative agent deliberation. It gives the basic concepts of roles and context and how these are captured through dynamic priorities in argumentation. Section 4 discusses related and future work.

2 The Argumentation Framework of LPwNF

An agent has his own theory expressing the knowledge under which he will take decisions. This decision process needs to compare alternatives and arrive at a conclusion that reflects a certain policy of the agent. In this paper we formalize this type of agent reasoning via argumentation where the deliberation of an agent is captured through an argumentative evaluation of arguments and counter-arguments.

There are several frameworks of argumentation proposed recently (e.g. [18,4]) that could be adopted for formalizing an agent's deliberation. We will use the framework presented in [9,6], called *Logic Programming without Negation as Failure* ($LPwNF$) (The historical reasons for this name are not directly relevant to this paper). We briefly review this framework and then study its extension needed to accommodate roles and context in argumentative deliberation.

In $LPwNF$ a non-monotonic argumentation theory is viewed as a pool of sentences (or rules) from which we must select a suitable subset, i.e. an argument, to reason with, e.g. to support a conclusion. Sentences in a $LPwNF$ theory are written in the usual extended logic programming language with an explicit negation, but without the Negation as Failure (NAF) operator. We will often refer to the sentences of a theory as argument rules. In addition, these rules may be assigned locally a "relative strength" through a partial ordering relation. For example, we may have

$fly(X) \leftarrow bird(X)$ $\qquad \neg fly(X) \leftarrow penguin(X)$
$bird(X) \leftarrow penguin(X)$ $\qquad bird(tweety)$

with an ordering relation between the rules that assigns the second rule higher than the first. This theory captures the usual example of "flying birds" with its exceptions, without the use of explicit qualifications of the default rules with abnormality conditions. We can conclude that *tweety* flies since we can derive this from the first rule and there is no way to derive $\neg fly(tweety)$. We have an argument (i.e. a proof) for $fly(tweety)$ but no argument for $\neg fly(tweety)$. If we add to the theory $penguin(tweety)$ then we can derive both $fly(tweety)$

and $\neg fly(tweety)$ - we have an argument for either conclusion. But in the non-monotonic argumentation semantics of the theory we can only conclude $\neg fly(tweety)$. This overrides $fly(tweety)$ since the argument that derives $\neg fly(tweety)$ contains the second rule which is designated higher than the first rule which belongs to the argument that derives $fly(tweety)$. We say that the argument for $\neg fly(tweety)$ **attacks** the argument for $fly(tweety)$ but not vice-versa. In general, the argumentation-based framework of $LPwNF$ is defined as follows.

Definition 2.1 Formulae in the **background logic** $(\mathcal{L}, \vdash)$ of the framework are defined as $L \leftarrow L_1, \ldots, L_n$, where $L, L_1, \ldots, L_n$ are positive or explicit negative ground literals. The derivability relation, $\vdash$, of the logic is given by the single inference rule of modus ponens [3].

Together with the set of sentences of a theory $\mathcal{T}$, we are given an ordering relation $<$ on these sentences (where $\phi < \psi$ or $<(\phi, \psi)$ means that ϕ has lower priority than ψ). The role of the priority relation is to encode locally the relative strength of argument rules in the theory. The relation $<$ is required to be irreflexive.

Definition 2.2 An **argumentation theory** $(\mathcal{T}, <)$ is a set of sentences $\mathcal{T}$ in $\mathcal{L}$ together with a priority relation $<$ on the sentences of $\mathcal{T}$. An **argument** for a literal L in a theory $(\mathcal{T}, <)$ is any subset of $\mathcal{T}$ that derives L, $T \vdash L$, under the background logic.

In general, we can separate out a part of the theory $\mathcal{T}_0 \subset \mathcal{T}$ (e.g. the last two rules of the example above) and consider this as a non-defeasible part from which any argument rule can draw information that it might need. The notion of attack between arguments in a theory $\mathcal{T}$ is based on the possible conflicts between a literal L and its explicit negation $\neg L$ and on the priority relation $<$ on $\mathcal{T}$.

Definition 2.3 Let $(\mathcal{T}, <)$ be a theory and $T, T' \subseteq \mathcal{T}$. Then T' **attacks** T (or T' is a **counter argument** of T) iff there exists L, $T_1 \subseteq T'$ and $T_2 \subseteq T$ s.t.:

(i) $T_1 \vdash_{min} L$ and $T_2 \vdash_{min} \neg L$

(ii) $(\exists r' \in T_1, r \in T_2 \text{ s.t. } r' < r) \Rightarrow (\exists r' \in T_1, r \in T_2 \text{ s.t. } r < r')$.

Here $T \vdash_{min} L$ means that $T \vdash L$ under the background logic and that L can not be derived from any proper subset of T. The second condition in this definition states that an argument T' for L attacks an argument T for the contrary conclusion only if the set of rules that it uses to prove L are at least of the same strength (under the priority relation $<$) as the set of rules in T

[3] The background logic of this argumentation framework can be replaced with any monotonic first order logic.

used to prove the contrary. Note that the attacking relation is not necessarily symmetric.

Using this notion of attack we then define the central notions of an *admissible argument* of a given theory and the non-monotonic argumentation consequence relation of a given theory as follows.

Definition 2.4 Let $(\mathcal{T}, <)$ be a theory and T a subset of $\mathcal{T}$. Then T is **admissible** iff T is consistent and for any $T' \subseteq \mathcal{T}$ if T' attacks T then T attacks T'.

Definition 2.5 Let $T = (\mathcal{T}, <)$ be a theory and L a ground literal. Then L is a **credulous (resp. skeptical) consequence** of T iff L holds in a (resp. every) maximal (wrt set inclusion) admissible subset of $\mathcal{T}$.

3 Argumentative Agent Deliberation

In this section we will extend the $LPwNF$ framework to allow dynamic priorities and use this to formulate the general framework of argumentative agent deliberation. Within this framework we will be interested in capturing the basic concepts of agent roles and context of interaction between agents.

Agents are always integrated within a (social) environment of interaction. We call this the *context* of interaction. This determines relationships between the possible roles the different agents can have within the environment. We consider, in line with much of the agent literature, (e.g. [16,23]), a *role* as a set of behaviour obligations, rights and privileges determining its interaction with other roles.

Generally, the substance of roles is associated to a *default context* that defines shared social relations of different forms (e.g. authority, friendship, relationship, etc.) and specifies the behaviour of roles between each others. Consequently, it implicitly installs a partial order between roles that expresses preferences of behaviour. For instance in the army context an officer gives orders that are obeyed by a soldier, or in a everyday context we respond in favour more easily to a friend than to a stranger. However, a default context that determines the basic roles filled by the agents is not the only environment where they could interact. For example, two friends can also be colleagues or an officer and a soldier can be family friends in civil life. Therefore we consider a second level of context, called *specific context*, which can overturn the pre-imposed, by the default context, ordering between roles and establish a different social relation between them. For instance, the authority relationship between an officer and a soldier would change under the specific context of a social meeting at home or the specific context of treason by the officer.

3.1 Argumentation with Roles and Context

In order to accommodate in an agent's argumentative reasoning the roles and context as described above we can extend the framework of $LPwNF$ so that the priority relation of a theory is not simply a static relation but a dynamic relation that captures the non-static preferences associated to roles and context. There is a natural way to do this. Following the same philosophy of approach as in [19], the priority relation can be defined as part of the agent's theory $\mathcal{T}$ and then be given the same argumentation semantics along with the rest of the theory.

We distinguish the part of the theory that defines the priority relation by $\mathcal{P}$. Rules in $\mathcal{P}$ have the same form as any other rule, namely ground rules of the form $L \leftarrow L_1, \ldots, L_n$ where the head L refers to the (irreflexive) higher-priority relation, i.e. L has the general form $L = h\_p(rule1, rule2)$. Also for any ground atom $h\_p(rule1, rule2)$ its negation is denoted by $h\_p(rule2, rule1)$ and vice-versa. For simplicity of presentation we will assume that the conditions of any rule in the theory do not refer to the predicate h_p thus avoiding self-reference problems. We now need to extend the semantic definitions of attack and admissibility.

Definition 3.1 Let $(\mathcal{T}, \mathcal{P})$ be a theory, $T, T' \subseteq \mathcal{T}$ and $P, P' \subseteq \mathcal{P}$. Then (T', P') **attacks** (T, P) iff there exists a literal L, $T_1 \subseteq T'$, $T_2 \subseteq T$, $P_1 \subseteq P'$ and $P_2 \subseteq P$ s.t.:

(i) $T_1 \cup P_1 \vdash_{min} L$ and $T_2 \cup P_2 \vdash_{min} \neg L$

(ii) $(\exists r' \in T_1 \cup P_1, r \in T_2 \cup P_2$ s.t. $T \cup P \vdash h\_p(r, r')) \Rightarrow (\exists r' \in T_1 \cup P_1, r \in T_2 \cup P_2$ s.t. $T' \cup P' \vdash h\_p(r', r))$.

Here, when L does not refer to h_p, $T \cup P \vdash_{min} L$ means that $T \vdash_{min} L$. This extended definition means that a composite argument (T', P') is a counter-argument to another such argument when they derive a contrary conclusion, L, and $(T' \cup P')$ makes the rules of its counter proof at least "as strong" as the rules for the proof by the argument that is under attack. Note that now the attack can occur on a contrary conclusion L that refers to the priority between rules.

Definition 3.2 Let $(\mathcal{T}, \mathcal{P})$ be a theory, $T \subseteq \mathcal{T}$ and $P \subseteq \mathcal{P}$. Then (T, P) is **admissible** iff $(T \cup P)$ is consistent and for any (T', P') if (T', P') attacks (T, P) then (T, P) attacks (T', P').

Hence when we have dynamic priorities, for an object-level argument (from $\mathcal{T}$) to be admissible it needs to take along with it priority arguments (from $\mathcal{P}$) to make itself at least "as strong" as the opposing counter-arguments. This need for priority rules can repeat itself when the initially chosen ones can themselves be attacked by opposing priority rules and again we would need to make now the priority rules themselves at least as strong as their opposing ones.

We can now define an agent's argumentation theory for describing his policy in an environment with roles and context as follows.

Definition 3.3 An agent's **argumentative policy theory or theory,** T, is a triple $T = (\mathcal{T}, \mathcal{P}_R, \mathcal{P}_C)$ where the rules in $\mathcal{T}$ do not refer to h_p, all the rules in $\mathcal{P}_R$ are priority rules with head $h\_p(r_1, r_2)$ s.t. $r_1, r_2 \in \mathcal{T}$ and all rules in $\mathcal{P}_C$ are priority rules with head $h\_p(R_1, R_2)$ s.t. $R_1, R_2 \in \mathcal{P}_R \cup \mathcal{P}_C$.

We therefore have three levels in an agent's theory. In the first level we have the rules $\mathcal{T}$ that refer directly to the subject domain of the agent. We call these the **Object-level Decision Rules** of the agent. In the other two levels we have rules that relate to the policy under which the agent uses his object-level decision rules according to roles and context. We call the rules in $\mathcal{P}_R$ and $\mathcal{P}_C$, **Role (or Default Context) Priorities** and **(Specific) Context Priorities** respectively.

As an example, consider the following theory $\mathcal{T}$ representing (part of) the object-level decision rules of an employee in a company [4].

$r_1(A, Obj, A_1) : give(A, Obj, A_1) \leftarrow requests(A_1, Obj, A)$
$r_2(A, Obj, A_1) : \neg give(A, Obj, A_1) \leftarrow needs(A, Obj)$
$r_3(A, Obj, A_2, A_1) : \neg give(A, Obj, A_2) \leftarrow give(A, Obj, A_1), A_2 \neq A_1.$

In addition, we have a theory $\mathcal{P}_R$ representing the general default behaviour of the code of contact in the company relating to the roles of its employees: a request from a superior is in general stronger than an employee's own need; a request from another employee from a competitor department is in general weaker than its own need. Here and below we will use capitals to name the priority rules but these are not to be read as variables. Also for clarity of presentation we do not write explicitly the full name of a priority rule omitting in the name the ground terms of the rules.

$R_1 : h\_p(r_1(A, Obj, A_1), r_2(A, Obj, A_1)) \leftarrow higher\_rank(A_1, A)$
$R_2 : h\_p(r_2(A, Obj, A_1), r_1(A, Obj, A_1)) \leftarrow competitor(A, A_1)$
$R_3 : h\_p(r_1(A, Obj, A_1), r_1(A, Obj, A_2)) \leftarrow higher\_rank(A_1, A_2)$

Between the two alternatives to satisfy a request from a superior from a competing department or not, the first is stronger when these two departments are in the specific context of working together on a common project. On the other hand, if we are in a case where the employee who has an object and needs it, needs this urgently then he would prefer to keep it. Such policy is represented at the third level in $\mathcal{P}_C$:

$C_1 : h\_p(R_1(A, Obj, A_1), R_2(A, Obj, A_1)) \leftarrow common(A, Obj, A_1)$
$C_2 : h\_p(R_2(A, Obj, A_1), R_1(A, Obj, A_1)) \leftarrow urgent(A, Obj).$

Note the *modularity* of this representation. For example, if the company decides to change its policy "that employees should generally satisfy the requests

[4] Non-ground rules represent their instances in a given Herbrand universe.

of their superiors" to apply only to the direct manager of an employee we would simply replace R_1 by the new rule R_1' without altering any other part of the theory:

$$R_1' : h\_p(r_1(A, Obj, A_1), r_2(A, Obj, A_1)) \leftarrow manager(A_1, A).$$

Consider now a scenario where we have two agents ag_1 and ag_2 working in competing departments and that ag_2 requests an object from ag_1. This is represented by extra statements in the non-defeasible part, $\mathcal{T}_0$, of the theory, e.g. $competitor(ag_2, ag_1)$, $requests(ag_2, obj, ag_1)$. Should ag_1 give the object to ag_2 or not?

If ag_1 does not need the object then, there are only admissible arguments for giving the object, e.g. $\Delta_1 = (\{r_1(ag_1, obj, ag_2)\}, \{\})$ and supersets of this. This is because this does not have any counter-argument as there are no arguments for not giving the object since $needs(ag_1, obj)$ does not hold. Suppose now that $needs(ag_1, obj)$ does hold. In this case we do have an argument for not giving the object, namely $\Delta_2 = (\{r_2(ag_1, obj, ag_2)\}, \{\})$. This is of the same strength as Δ_1 but the argument Δ_2', formed by replacing in Δ_2 its empty set of rules of priority with $\{R_2(r_2(ag_1, obj, ag_2), r_1(ag_1, obj, ag_2))\}$, attacks Δ_1 and any of its supersets but not vice-versa: R_2 gives higher priority to the rules of Δ_2 and there is no set of priority rules with which we can extend Δ_1 to give its object-level rules equal priority as those of Δ_2. Hence we conclude skeptically that ag_1 will not give the object. This skeptical conclusion was based on the fact that the theory of ag_1 cannot prove that ag_2 is of higher rank than himself. If the agent learns that $higher\_rank(ag_2, ag_1)$ does hold then Δ_2' and Δ_1', obtained by adding to the priority rules of Δ_1 the set $\{R_1(r_1(ag_1, obj, ag_2), r_2(ag_1, obj, ag_2))\}$, attack each other. Each one of these is an admissible argument for not giving or giving the object respectively and so we can draw both conclusions credulously.

Suppose that we also know that the requested object is for a common project of ag_1 and ag_2. The argument Δ_2' is now not admissible since now it has another attack obtained by adding to the priority rule of Δ_1' the extra priority rule $C_1(R_1(ag_1, obj, ag_2), R_2(ag_1, obj, ag_2))$ thus strengthening its derivation of $h\_p(r_1, r_2)$. The attack now is on the contrary conclusion $h\_p(r_1, r_2)$. In other words, the argumentative deliberation of the agent has moved one level up to examine what priority would the different roles have, within the specific context of a common project. Δ_2' cannot attack back this attack and no extension of it exists that would strengthen its rules to do so. Hence there are no admissible arguments for not giving and ag_1 draws the skeptical conclusion to give the object.

We have seen in the above example that in several cases the admissibility of an argument depends on whether we have or not some background information about the specific case in which we are reasoning. For example, ag_1 may not have information on whether their two departments are in competition or not. This means that ag_1 cannot build an admissible argument for not giving the object as he cannot use the priority rule R_2 that he might like to do. But this information maybe just unknown and if ag_1 wants to find a way to refuse the request he can reason further to find *assumptions* related to the unknown information under which he can build an admissible argument. Hence in this example he would build an argument for not giving the object to ag_2 that is *conditional* on the fact that they belong to competing departments. Furthermore, this type of information may itself be dynamic and change while the rest of the theory of the agent remains fixed, e.g. ag_1 may have in his theory that ag_2 belongs to a competing department but he has not yet learned that ag_2 has changed department or that his department is no longer a competing one.

We can formalize this conditional form of argumentative reasoning by defining the notion of *supporting information* and extending argumentation with *abduction* on this missing information.

Definition 3.4 Let $T = (\mathcal{T}_0, \mathcal{T}, \mathcal{P})$ be a theory, and $\mathcal{A}$ a distinguished set of predicates in the language of the theory, called **abducible** predicates [5]. Given a goal G, a set S of abducible literals consistent with the non-defeasible part $\mathcal{T}_0$ of T, is called a **strong (resp. weak) supporting evidence** for G iff G is a skeptical (resp. credulous) consequence of $(\mathcal{T}_0 \cup S, \mathcal{T}, \mathcal{P})$.

The structure of an argument can also be generalized as follows.

Definition 3.5 Let $T = (\mathcal{T}_0, \mathcal{T}, \mathcal{P})$ be a theory and $\mathcal{A}$ its abducible predicates. A **supported argument** in T is a tuple (Δ, S), where S is a set of abducible literals consistent with $\mathcal{T}_0$ and Δ is a set of argument rules in T, which is not admissible in T, but is admissible in the theory $(\mathcal{T}_0 \cup S, \mathcal{T}, \mathcal{P})$. We say that S supports the argument Δ.

The supporting information expressed through the abducibles predicates refers to the incomplete and evolving information of the external environment of interaction. Typically, this information pertains to the context of the environment, the roles between agents or any other aspect of the environment that is dynamic. We will see below an example of how agents can acquire and/or validate such information through a scenario of interaction where they exchange missing information.

Given the above framework the **argumentative deliberation** of an agent can be formalized via the following basic reasoning functions.

Definition 3.6 Let Ag be an agent, T his argumentation theory, G a goal and S a set of supporting information consistent with $\mathcal{T}_0$. Then we say that Ag **deliberates** on G, supported by S, to produce s^{ag}, denoted by $deliberate(Ag, G, S; s^{ag})$, iff $s^{ag} \neq \{\}$ is a strong supporting evidence for G in the theory $T \cup S$. If $s^{ag} = \{\}$ then we say that Ag *accepts* G under

[5] Typically, the theory $\mathcal{T}$ does not contain any rules for the abducible predicates.

$T \cup S$ and is denoted by **accept(Ag, G, S)**. Furthermore, given an opposing goal $\overline{G}$ (e.g $\neg G$) to G and s' produced by deliberation on $\overline{G}$, i.e. that $deliberate(Ag, \overline{G}, S; s')$ holds, we say that s' is supporting evidence for agent Ag to **refuse** G in $T \cup S$.

We will now illustrate the use of the argumentative deliberation of an agent, defined above, within a simple interaction protocol where two agents are trying to agree on some goal, as an example of how this argumentation framework can be used within the different decision making processes of an agent. In our study of this we will be mainly interested how agents can use their argumentative deliberation in order to decide their position at each step of the interaction process. We will not be concerned with the conversation protocol supporting the agent interaction.

Let us therefore consider a scenario where two self-interested agents, a seller and a buyer, are trying to agree on a goal to determine the price of a certain product. In particular, let us consider a seller agent called X who has the goal, G^X, to sell a product at a high price to another agent, the buyer, called Y, who has the (opposing) goal, G^Y, to buy this product at a low price. They are trying to find an agreement on the price by agreeing either on G^X or on G^Y. Furthermore, we will assume that during their interaction each agent initially insists on its own goal exchanging between them supporting information for their respective goals. When one of the agents cannot find anymore, within his theory and the accumulated supporting information (agreed by the two agents so far), an argument for his own goal he then considers the goal of the other agent. He deliberates on this to see if he can find, again according to his own theory and the supporting information agreed on so far, an argument that would support this goal of the other agent.

We assume that the seller has the following argumentation policy for selling products. We present only a part of this theory.
The object-level theory $\mathcal{T}^X$ of the seller contains the rules:

$r_1 : sell(Prd, A, high\_price) \leftarrow pay\_normal(A, Prd)$
$r_2 : sell(Prd, A, high\_price) \leftarrow pay\_install(A, Prd)$
$r_3 : sell(Prd, A, low\_price) \leftarrow pay\_cash(A, Prd)$
$r_4 : \neg sell(Prd, A, P_2) \leftarrow sell(Prd, A, P_1), P_2 \neq P_1.$

His role and context priority theories, $\mathcal{P}_R^X$ and $\mathcal{P}_C^X$, are given below. They contain the policy of the seller on how to treat the various types of customers. For example, to prefer to sell with normal paying conditions over payment by installments when the buyer is a normal customer (see R_1). Also that there is always a preference to sell at high price (see R_2, R_3) but for regular customers there are conditions under which the seller would sell at low price (see R_4, R_5). This low price offer to a regular customer applies only when we are not in high season (see C_1, C_2).

$R_1 : h\_p(r_1(Prd, A), r_2(Prd, A)) \leftarrow normal(A)$
$R_2 : h\_p(r_1(Prd, A), r_3(Prd, A))$
$R_3 : h\_p(r_2(Prd, A), r_3(Prd, A))$
$R_4 : h\_p(r_3(Prd, A), r_1(Prd, A)) \leftarrow regular(A), buy\_2(A, Prd)$
$R_5 : h\_p(r_3(Prd, A), r_1(Prd, A)) \leftarrow regular(A), late\_del(A, Prd)$
$C_1 : h\_p(R_2(Prd, A), R_4(Prd, A)) \leftarrow high\_season$
$C_2 : h\_p(R_2(Prd, A), R_5(Prd, A)) \leftarrow high\_season$
$C_3 : h\_p(R_4(Prd, A), R_5(Prd, A))$.

Given this theory the seller will initially offer to sell at a high price under normal paying conditions assuming that the buyer is a normal customer. Let us suppose that the buyer responds that this price is expensive. Given this new information the seller agent deliberating again on his goal, finds another argument for selling high, using now the object-level rule r_2 since he can no longer consider the buyer a normal customer and hence R_1 does not apply (the seller derives this from some general background knowledge that he has in $\mathcal{T}_0$ e.g. from a rule $\neg normal(A) \leftarrow expensive(A, high\_price)$). This new argument needs the support $pay\_install(buyer, prd)$ and the seller offers this information to the buyer.
Suppose that the buyer refuses this method of payment. Then the seller agent does not have any other argument (that is allowed given the information exchanged so far) to support his own goal of selling high. He could then consider the goal of the buyer, i.e. to sell at *low_price*. Deliberating on this goal he finds that his argumentation theory can support this if the customer is a regular one and he can accept some other conditions. He indeed finds an admissible argument for low price using the object-level rule r_3 and the role priority rule R_4. This is conditional on the information that the buyer is indeed a regular customer, will pay cash and that he will buy two of the products. Note that for this argument to be admissible the context rule C_1 must not apply, i.e. the seller knows that currently they are not in a *high_season*. Suppose that the buyer confirms the first two conditions but refuses the third. The seller then has another argument for selling low to a regular customer conditional on late delivery.

3.2 Modularity and Computation

As mentioned above, the proposed framework allows modular representations of problems where a change in the policy of an agent can be effected locally in his theory. The following results formalize some of the properties of modularity of the framework.

Proposition 3.7 *Let Δ be a set of arguments that is admissible separately with respect to the theory $T_1 = (\mathcal{T}, \mathcal{P}_{R1}, \{\})$ and the theory $T_2 = (\mathcal{T}, \mathcal{P}_{R2}, \{\})$. Then Δ is admissible with respect to the theory $T = (\mathcal{T}, \mathcal{P}_{R1} \cup \mathcal{P}_{R2}, \{\})$. Sim-*

ilarly, we can decompose $\mathcal{P}_C$ into $\mathcal{P}_{C1}$ and $\mathcal{P}_{C2}$.

Proposition 3.8 *Let Δ be a set of arguments that is admissible with respect to the theory $T_1 = (\mathcal{T}, \mathcal{P}_R, \{\})$. Suppose also that Δ is admissible with respect to $T_2 = (\mathcal{T} \cup \mathcal{P}_R, \{\}, \mathcal{P}_C)$. Then Δ is admissible with respect to $T = (\mathcal{T}, \mathcal{P}_R, \mathcal{P}_C)$.*

The later proposition shows that we can build an admissible argument $\Delta = (O, R)$ by joining together an object-level argument O together with a set of priority rules R that makes O admissible and is itself admissible with respect to the higher level of context priorities. These results provide the basis for a modular computational model in terms of interleaving levels of admissibility processes one for each level of arguments in the theory.

In general, the basic $LPwNF$ has a simple and well understood computational model [6] that can be seen as a realization of a more abstract computational model for argumentation [13]. It has been successfully used [12] to provide a computational basis for reasoning about actions and change. The simple argumentation semantics of $LPwNF$, where the attacking relation between arguments depends only on the priority of the rules of a theory, gives us a natural "dialectical" proof theory for the framework. In this we have two types of interleaving derivations one for considering the attacks and one for counter attacking these attacks. The proof theory then builds an admissible argument for a given goal by incrementally considering all its attacks and, whenever an attack is not counter-attacked by the argument that we have build so far, we extend this with other arguments (rules) so that it does so. This in turn may introduce new attacks against it and the process is repeated.

The priorities amongst the rules help us move from one type of derivation to the other type e.g. we need only consider attacks that come from rules with strictly higher priority than the rules in the argument that we are building (as otherwise the argument that we have so far will suffice to counter attack these attacks.) For the more general framework with dynamic priorities we apply the same proof theory extended and a derivation is thus split into levels. Now a potential attack can be avoided by ensuring that its rules are not of higher priority than the argument rules we are building and hence we move the computation one level up to attacks and counter attacks on the priorities of rules. This move one level can then be repeated to bring us to a third level of computation.

This extended proof theory has been implemented and used to build agents that deliberate in the face of complete (relevant) information of their environments. We are currently investigating how to extend this implementation further with (simple forms of ground) abduction, required for the computation of supporting evidence in the face of incomplete information about the environment, using standard methods from abductive logic programming.

4 Related Work and Conclusions

In this paper we have proposed an argumentative deliberation framework for autonomous agents and presented how this could be applied. We have argued that this framework has various desired properties of simplicity and modularity and in particular we have shown how it can capture some natural aspects of the decision making process of an autonomous agent. The framework can embody in a direct and modular way a preference policy of the agent which can be used to exhibit natural behaviour in an argumentation based interaction with other agents. The proposed argumentation framework has a simple and modular computational model that can be used to implement easily deliberative agents.

The argumentation framework developed and used in this paper is based on the more general and abstract notions that have emerged from a series of previous studies on argumentation [11,8,10,7,9]. The basic notion that is used is that of admissibility [7] which is itself a special case of acceptability [9]. It also follows the more recent approach of [19,5] who have shown the need for dynamic priorities within argumentation when we want to apply this to formalize law and other related problems. Our framework is close to that of [19] in that it uses a similar background language of logic programming. They also both have a computational model that follows a dialectical pattern in terms of interleaving processes one for each level of arguments in the theory. In comparison our framework is simpler using only a single notion of attack and avoiding the separate use of negation as failure that it is subsumed by the use of rule priorities. In [5] dynamic priorities are related to the argumentation protocols, also called rules of order, describing which speech acts are legal in a particular state of the argumentation. Although the interests for application of our argumentation framework are different the formal relation to these frameworks is an interesting problem for further study.

In the development of agent deliberation we have introduced, in the same spirit as [21,2], roles and context as a means to define non-static priorities between arguments of an agent. This helps to capture the social dimension of agents, as it incorporates in a natural way the influence of the environment of interaction (which includes other agents) on the agents "way of thinking and acting". We have shown how we can encompass, within this framework, the relative roles of agents and how these can vary dynamically depending on the external environment. The representation of this role and context information is expressed directly in terms of priority rules which themselves form arguments in the framework and are reasoned about in the same way as the object level arguments. As a result this gives a high-level encapsulation of these notions where changes are easily accommodated in a modular way.

The use of roles and dynamic context is a basic difference with most of other works [22,17,15,3,1] on agent argumentation. Our work complements and extends the approaches of [21,2] with emphasis on enriching the self argumentative deliberation of an agent. It complements these works by linking directly the preferences between different contexts, which these works propose, to a first level of roles that agents can have in a social context, called default context, showing how roles can be used to define in a natural way priorities between arguments of the agents filling these roles. It extends this previous work by incorporating reasoning on these preferences within the process of argumentative deliberation of an agent. This is done by introducing another dimension of context, called specific context, corresponding to a second level of deliberation for the agent. This allows a higher degree of flexibility in the adaptation of the agents argumentative reasoning to a dynamically changing environment. In [2] the context preferences can also be dynamic but the account of this change is envisaged to occur outside the argumentative deliberation of the agent. An agent decides a-priori to change the context in which he is going to deliberate. In our case the change is integrated within the deliberation process of the agent.

This extra level of deliberation allows us to capture the fact that recognized roles in a context have their impact only within the default context where they are defined, although these roles always "follow" agents filling them, as a second identity in any other context they find themselves. Therefore agents who have some relationship, imposed by their respective roles, can be found in a specific context where the predefined (according to their relationship) order of importance between them has changed.

In comparison with other works on agent argumentation our work also integrates abduction with argumentation to handle situations where the information about the environment, currently available to the agent, is incomplete. This use of abduction is only of a simple form and more work is needed to study more advanced uses of abduction drawing from recent work on abduction in agents [20]. Another direction of future work concerns dialogue modeling. Our aim is to use our argumentative deliberation model for determining dialogue acts and protocols by extending the dialogue framework presented in [14]. Currently, we are also studying how to use the proposed argumentation framework together with work from cognitive psychology to model needs and motivations of agents and through this to define, via argumentation theories, different agent personalities.

References

[1] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *ICMAS-00, pp. 31-38*, 2000.

[2] L. Amgoud and S. Parsons. Agent dialogues with conflicting preferences. In *ATAL01*, 2001.

[3] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue and negotiation. In *ECAI-00, pp. 338-342*, 2000.

[4] A. Bondarenko, P. M. Dung, R. A. Kowalski, and F. Toni. An abstract, argumentation-theoretic framework for default reasoning. *Artificial Inelligence*, 93(1-2):63–101, 1997.

[5] G. Brewka. Dynamic argument systems: a formal model of argumentation process based on situation calculus. In *Journal of Logic and Computation, 11(2), pp. 257-282*, 2001.

[6] Y. Dimopoulos and A. C. Kakas. Logic programming without negation as failure. In *Proc. ILPS'95, pp. 369-384*, 1995.

[7] P.M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. In *Artificial Intelligence, 77, pp. 321-357 (also in IJCAI'93)*, 1995.

[8] P.M. Dung, A. C. Kakas, and P. Mancarella. Negation as failure revisited. In *University of Pisa Technical Report*, 1992.

[9] A. C. Kakas, P. Mancarella, and P.M. Dung. The acceptability semantics for logic programs. In *Proc. ICLP'94, pp. 504-519*, 1994.

[10] A.C. Kakas. Default reasoning via negation as failure. In *LNAI, Vol. 810, pp. 160-179*, 1994.

[11] A.C. Kakas, R.A. Kowalski, and F. Toni. Abductive logic programming. In *Journal of Logic and Computation, 2(6), pp. 719-770*, 1992.

[12] A.C. Kakas, R.S. Miller, and F. Toni. E-res: Reasoning about actions, events and observations. In *LPNMR'01, pp. 254-266*, 2001.

[13] A.C. Kakas and F. Toni. Computing argumentation in logic programming. In *JLC 9(4), 515–562, O.U.P*, 1999.

[14] N. Karacapilidis and P. Moraitis. Engineering issues in inter-agent dialogues. In *Proc. of 15th European Conference on Artificial Intelligent (ECAI02), Lyon, France,*, 2002.

[15] S. Kraus, K. Sycara, and A. Evenchik. Reaching agreements through argumentation: a logical model and implementation. In *Artificial Intelliegence, 104 pp. 1-69*, 1998.

[16] P. Panzarasa, N.R. Jennings, and T. Norman. Formalising collaborative decision-making and practical reasoning in multi-agent systems. In *Journal of Logic and Computation 12 (1), to appear*, 2002.

[17] S. Parsons, C. Sierra, and N.R. Jennings. Agents that reason and negotiate by arguying. In *Logic and Computation 8 (3), 261-292*, 1998.

[18] J.L. Pollock. Justification and defeat. In *Artficial Intelligence Vol 67, pp. 377-407*, 1994.

[19] H. Prakken and G. Sartor. A dialectical model of assessing conflicting arguments in legal reasoning. In *Artficial Intelligence and Law Vol 4, pp. 331-368*, 1996.

[20] F. Sadri, F. Toni, and P. Torroni. Dialogues for negotiation: agent varieties and dialogue sequences. In *ATAL01*, 2001.

[21] C. Sierra, N.R. Jennings, P. Noriega, and S. Parsons. A framework for argumentation-based negotiation. In *ATAL-97, pp. 167-182*, 1997.

[22] K. Sycara. Argumentation: Planning other agents' plans. In *IJCAI-89, pp. 517-523*, 1989.

[23] M. Wooldridge, N.R. Jennings, and D. Kinny. The gaia methodology for agent-oriented analysis and design. In *JAAMAS 3 (3), pp. 285-312*, 2000.