Autonomous Agents and Multi-Agent Systems, 11, 189–206, 2005

Modular Representation of Agent Interaction Rules through Argumentation

ANTONIS KAKAS antonis@cs.ucy.aac.cy
Department of Computer Science, University of Cyprus, CY-1678 Nicosia, Cyprus

NICOLAS MAUDET maudet@lamsade.dauphine.fr
LAMSADE, Univ. Paris-Dauphine, 75775 Paris Cedex 16, France

PAVLOS MORAITIS moraitis@cs.ucy.ac.cy
Department of Computer Science, University of Cyprus, CY-1678 Nicosia, Cyprus

Abstract. Communication between agents needs to be flexible enough to encompass together a variety of different aspects such as, conformance to society protocols, private tactics of the individual agents, strategies that reflect different classes of agent types (or personal attitudes) and adaptability to the particular external circumstances at the time when the communication takes place. In this paper, we propose an argument-based framework for representing communication theories of agents that can take into account in a uniform way these different aspects. We show how this approach can be used to realize existing types of dialogue strategies and society protocols in a way that facilitates their modular development and extension to make them more flexible in handling different or special circumstances.

Keywords: multi-agent systems, argumentation, protocols, strategies.

1. Introduction

Communication is one of the main features of multiagent systems. Society *protocols* regulate the communicative behaviour agents should conform to by defining what dialogue moves are legal in any given situation. Private *strategies*, as adopted by an individual agent, specify the dialogue move(s) the agent is willing to utter, according to its own objectives and other personal characteristics. Ideally, dialogue moves selected by the agent's strategy will fall within the legal moves defined by the protocol.

In this paper, we investigate how to represent communication patterns using an argumentation-based framework with dynamic preferences. The behaviour of an agent participating in a dialogue is conditioned on two theories in this framework each one of which expressed as a preference policy on the dialogue moves.

The first theory captures the society protocol describing the legal moves at two levels, normal (or default) and exceptional, as the preferred moves given a current set of circumstances within the society the agent belongs to. The context-dependent protocols afforded by our representation framework will give a high degree of flexibility to encompass, together in one uniform theory, the different aspects of the protocol under different circumstances, as perceived by the different agents in the society that share this common protocol.

The second theory describes, again as a preference policy, the personal strategy of the agent. This can be influenced by application domain tactics but also by the agent's personal *profile or attitude* characteristics. As with the society protocols, this theory is context-dependent in order to take into account the variety of situations under which a dialogue can take place (e.g. the different *roles* of the interlocutors and the *context* of the dialogue) as well as the dynamically changing circumstances of the dialogue. The overall decision of which move to utter next is based on the integration of these theories by suitably exploiting the sceptical and credulous forms of argumentation-based reasoning. Our approach therefore allows the modular separation of concerns: professional tactics of dialogue strategy, personal attitudes influencing the strategy and legality of strategy decisions required by societal protocols.

Several works have studied the problem of dialogue strategies in interactions governed by social protocols, many of which [1, 13] use, like we do, argumentation as their basis. Our work can be viewed as providing an approach where these notions can be modularly realized and in cases extended to allow a wider class of problems to be addressed. This stems from the flexibility provided to define private strategies and public protocols uniformly within the same highly expressive representation framework, which in addition possesses a viable computational model. Communication theories can be easily implemented directly from their declarative specification in the Gorgias system [6] this framework.

The remainder of this paper is as follows. Section 2 gives the background framework on agent argumentative reasoning used in this paper. Section 3 studies the representation of agent private strategies while Section 4 explores in turn the representation of social protocols in the same framework. Section 5 studies in some detail the connection to existing approaches.

2. The agent reasoning framework

This section gives the basic concepts of the underlying argumentation framework in which an agent represents and reasons with its communication theory. This framework was proposed in Kakas et al. [8] and developed further in Kakas and Moraitis [9], in order to accommodate a dynamic notion of priority over the rules (and hence the arguments) of a given theory, as also advocated in Brewka [2] and Prakken and Sartor [14] for instance.

As proposed in McBurney and Parsons [11] we can distinguish three languages in the representation of an agent's communication theory. A language, $\mathcal{L}$, to describe the background information that the agent has about its world at any moment and the basic rules for deciding its communication moves; a language, $\mathcal{ML}$, for expressing preference policies pertaining to its decision of these moves; and a language, $\mathcal{CL}$, which is a common communication language for all agents.

Furthermore, we will see that (components of) an agent's theory will be layered in three levels. *Object-level decision rules*, in the language $\mathcal{L}$, are defined at the first level. The next two levels, represented in the language $\mathcal{ML}$, describe priority rules on the decision rules of the first level and on themselves thus expressing a preference policy for the overall decision making of the agent. This policy is separated into two levels:

level two to capture the *default* preference policy under normal circumstances while level three is concerned with the *exceptional* part of the policy that applies under specific contexts. Hence we will assume that agents are always associated with a (social) environment of interaction in which they can distinguish normal (or default) contexts from specific (or exceptional) contexts. Their argumentation-based decision making will then be sensitive to context changes.

In general, an argumentation theory is defined as follows.

Definition 2.1. *A theory is a pair* $(\mathcal{T}, \mathcal{P})$. *The sentences in* $\mathcal{T}$ *are formulae, in a background monotonic logic* $(\mathcal{L}, \vdash)$, *of the form* $L \leftarrow L_1, \ldots, L_n$, *where* $L, L_1, \ldots, L_n$ *are positive or explicit negative ground literals.* $\mathcal{P}$ *is a set of rules defined in the language* $\mathcal{ML}$ *which is the same as* $\mathcal{L}$ *apart from the fact that the head L of the rules has the general form* $L = h\_p(rule1, rule2)$ *where rule1 and rule2 are ground functional terms that name any two rules in the theory. This higher-priority relation given by h_p is required to be irreflexive. The derivability relation,* $\vdash$, *of the background logic for* $\mathcal{L}$ *and* $\mathcal{ML}$ *is given by the single inference rule of modus ponens.*

For simplicity, it is assumed that the conditions of any rule in the theory do not refer to the predicate *h_p* thus avoiding self-reference problems. For any ground atom $h\_p(rule1, rule2)$ its negation is denoted by $h\_p(rule2, rule1)$ and vice-versa.

An *argument* for a literal L in a theory $(\mathcal{T}, \mathcal{P})$ is any subset, T, of this theory that derives L, i.e. $T \vdash L$ under the background logic. The subset of rules in the argument T that belong to $\mathcal{T}$ is called the *object-level* argument. Note that in general, we can separate out a part of the theory $\mathcal{T}_0 \subset \mathcal{T}$ and consider this as a non-defeasible part from which any argument rule can draw information that it might need. We call $\mathcal{T}_0$ the background knowledge base.

The notion of attack between arguments in a theory is based on the possible conflicts between a literal L and its negation and on the priority relation of *h_p* in the theory.

Definition 2.2. *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 $S \vdash_{\min} L$ means that $S \vdash L$ and that no proper subset of S implies L. When L does not refer to *h_p*, $T \cup P \vdash_{\min} L$ means that $T \vdash_{\min} L$. This definition states that a "composite" argument (T', P') is a counter-argument to another such argument when it derives 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 the attack can occur on a contrary conclusion $L = h\_p(r, r')$ that refers to the priority between rules.

Definition 2.3. *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'). *Given a ground literal L then L is a credulous (respectively skeptical) consequence of the theory iff L holds in a (respectively every) maximal (wrt set inclusion) admissible subset of* $\mathcal{T}$.

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.

An agent's argumentation theory will be defined as a theory $(\mathcal{T}, \mathcal{P})$ which is further layered in separating $\mathcal{P}$ into two parts as follows.

Definition 2.4. *An agent's argumentative policy theory is a theory* $T = (\mathcal{T}, (\mathcal{P}_R, \mathcal{P}_C))$ *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 theory at hand. 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 its object-level decision rules, associated to normal situations (related to a default context) and specific situations (related to specific or exceptional contexts). We call the rules in $\mathcal{P}_R$ and $\mathcal{P}_C$, *Default* or *Normal Context Priorities* and *Specific Context Priorities* respectively.

We note that in representing an argumentation theory we will use non-ground rules as *schemas* of rules. These rules will be written in Logic Programming notation where any term begining with a capital letter denotes a variable and where all variables in a rule are universally quantified. Such a (non-ground) schema rule then represents all its ground instances in a given Herbrand universe. As an example of an agent's argumentation theory, consider the following theory $\mathcal{T}$ representing (part of) the object-level decision rules of an employee in a company.

$$
\begin{array}{llcl}
r_1(A, Obj, A_1): & give(A, Obj, A_1) & \leftarrow & requests(A_1, Obj, A) \\
r_2(A, Obj, A_1): & \neg give(A, Obj, A_1) & \leftarrow & needs(A, Obj) \\
r_3(A, Obj, A_2, A_1): & \neg give(A, Obj, A_2) & \leftarrow & give(A, Obj, A_1), A_2 \neq A_1
\end{array}
$$

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.

$$
\begin{array}{lll}
R_1: & h\_p(r_1(A, Obj, A_1), r_2(A, Obj, A_1)) & \leftarrow & higher\_rank(A_1, A) \\
R_2: & h\_p(r_2(A, Obj, A_1), r_1(A, Obj, A_1)) & \leftarrow & competitor(A, A_1) \\
R_3: & h\_p(r_1(A, Obj, A_1), r_1(A, Obj, A_2)) & \leftarrow & higher\_rank(A_1, A_2)
\end{array}
$$

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 needs it urgently, then he would prefer to keep it. Such policy is represented at the third level in $\mathcal{P}_C$:

$$
\begin{array}{lll}
C_1: & h\_p(R_1(A, Obj, A_1), R_2(A, Obj, A_1)) & \leftarrow & common(A, Obj, A_1) \\
C_2: & h\_p(R_2(A, Obj, A_1), R_1(A, Obj, A_1)) & \leftarrow & urgent(A, Obj)
\end{array}
$$

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

$$
R'_1: \quad h\_p(r_1(A, Obj, A_1), r_2(A, Obj, A_1)) \quad \leftarrow \quad manager(A_1, A)
$$

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 to be used for a common project involving both ag_1 and ag_2. The argument Δ'_2 is now not admissible since 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 *obj* to ag_2, and ag_1 draws the skeptical conclusion to give the object.

2.1. *The communication framework*

We assume that agents interact using *dialogue moves* or *locutions*. Once performed, these dialogue moves are added directly, or via some sort of public structure like *commitment stores* [7], to the agent background knowledge $\mathcal{T}_0$. That is,we assume that all these moves are perfectly perceived by the agents of the society. As detailed in Karacapilidis and Moraitis [10], the shared communication language of the agents, $\mathcal{CL}$, contains a set of communication performatives of the form $P(X, Y, S)$, where

– P is a performative type belonging to the set *Perf*;
– X and Y are the sender and the receiver of the performative, respectively;
– S is the content (i.e., body) of the performative;

S can contain elements (facts, rules, etc.) expressing arguments supporting the message. We shall not give further details here: as we are primarily concerned with argumentation policies specifying how agents decide upon the next move to utter, it is sufficient to identify the relevant parameters (P, X, Y and S) that can influence the definition of these policies. For the sake of presentation, we also omit the utterance time parameter.

The set of *performative types* (*Perf*) of the chosen communication language will also play a significant role. We may take this to be one of the current standards, as proposed for instance by the FIPA consortium [5]. However, as these standards do not include moves devoted to argumentation, as discussed in McBurney et al. [12], we shall use instead a set of performative types more suited to our purpose, inspired by those introduced in Amgoud and Parsons [1] and Sadri et al. [16].

$$Perf = \{request, propose, accept, refuse, challenge\}$$

In what follows, this set will also be used as a *label set*. As mentioned above we will assume that both the society and the agents interacting in the society share the same set of performative types *Perf*.

3. Flexible agent strategies

Based on the argumentation framework described in the previous section we can compose a *private or personal strategy* theory of an agent in three parts which modularly capture different concerns of the problem. These parts are:

– the *basic component*, T_{basic}, that defines the private *dialogue steps* of the dialogue
– the *tactical component*, $T_{tactical}$, that defines a private preference policy of (professional) tactics
– the *attitude component*, $T_{attitude}$, that captures general (application independent) characteristics of personal strategy of the agent type

We call $T_{basic} \cup T_{tactical}$ the *tactical theory* and $T_{basic} \cup T_{attitude}$ the *attitude theory*. Let us examine in turn these different components.

The basic component (T_{basic}). This component contains object-level rules in the language $\mathcal{L}$, defining the private *dialogue steps*, and are (for an agent X) of the form:

$$r_{j,i}(Y, S', S): \quad p_j(X, Y, S') \quad \leftarrow \quad p_i(Y, X, S), c_{ij}$$

where i, j belong to the label set *Perf* and c_{ij} (which can be empty) are called the *enabling conditions* of the dialogue step from the performative p_i to p_j. In other words, these are the conditions under which the agent X (whose theory this is) may utter p_j upon receiving p_i from agent Y. These conditions thus correspond to the rationality rules of [1] or the conditions of the dialogue constraints of [16]. As mentioned before, these rules and their names, $r_{j,i}(Y, S', S)$, are written in Logic Programming schema of rules representing compactly all rules obtained by grounding these over the Herbrand universe of the theory.

For simplicity, we will assume that the enabling conditions are evaluated in the non-defeasible part, $\mathcal{T}_0$, of the theory containing the background knowledge that the agent X has about the world and the dialogue so far. This essentially simplifies the attacking relation of the argumentation but this is not a significant simplification for the purposes of the work of this paper. The background knowledge base $\mathcal{T}_0$ also contains the rules:

$$\begin{array}{lcl} \neg p_j(X, Y, S) & \leftarrow & p_i(X, Y, S), i \neq j \\ \neg p_i(X, Y, S') & \leftarrow & p_i(X, Y, S), S' \neq S \end{array}$$

for every i and j in *Perf* and every subject S', S to express the general requirement that two different utterances are incompatible with each other.

This means that any argument for one specific utterance is potentially (depending on the priority rules in the other parts of the theory) an attack for any other different one. Hence any admissible set of arguments cannot contain rules that derive more than one utterance. In fact, with the basic component alone the theory can (easily) have several credulous conclusions for which could be the next utterance as the following example illustrates.

Example 1. *Let us consider an agent Bob equipped with a basic component containing the following simplified rules.*

$$
\begin{array}{llll}
r_{acc,req}(Y,P) & : accept(X,Y,P) & \leftarrow & request(Y,X,P), \\
 & & & have(X,P) \\
r_{ref,req}(Y,P) & : refuse(X,Y,P) & \leftarrow & request(Y,X,P) \\
r_{chall,req}(Y,P) & : challenge(X,Y,P) & \leftarrow & request(Y,X,P) \\
r_{prop,req}(Y,P,Q) & : propose(X,Y,Q) & \leftarrow & request(Y,X,P), \\
 & & & altern(P,Q)
\end{array}
$$

Now assume that Bob has received the dialogue move request$(Al, Bob, nail)$ *and that Bob currently holds a nail, i.e.*

$$\mathcal{T}_0 = \{request(Al, Bob, nail), have(Bob, nail)\}$$

Then the possible reply moves would be

$$accept(Bob, Al, nail), refuse(Bob, Al, nail), challenge(Bob, Al, nail)$$

corresponding to the different credulous consequences of the theory, with no further information to discriminate them. Note that if $\mathcal{T}_0$ *also contained altern* $(Bob, nail, hook)$*, then we would have propose*$(Bob, Al, hook)$ *as an additional credulous conclusion.*

The extra information needed to discriminate between these equally possible moves will typically come from the preference policies described in the other two components (tactical and attitude) of the private strategy theory.

The tactical component ($T_{tactical}$). This component defines a private preference policy that captures the professional tactics of the agent for how to decide amongst the alternatives enabled by the basic part of the theory. It consists of two sets $\mathcal{P}_R, \mathcal{P}_C$ of priority rules, written in the language $\mathcal{ML}$, at the two higher levels as defined in Section 2.

The rules in $\mathcal{P}_R$ express priorities over the dialogue step rules in the basic part. A simple pattern that one can follow in writting these rules is to consider the dialogue steps that refer to the same incoming move $p_i(Y, X, S)$ and then have rules of the following form:

$R^i_{k\|j}$:	$h\_p(r_{k,i}, r_{j,i})$	$\leftarrow$	$true$
$R^i_{j\|k}$:	$h\_p(r_{j,i}, r_{k,i})$	$\leftarrow$	SC_{jk}

where SC_{jk} are specific conditions that are evaluated in the background knowledge base of the agent and could depend on the agent Y, the content of the incoming move and indeed the types j and k of these alternative moves. Note that these rules R can have additional superscripts in their names if there is a need to distinguish them further.

The first rule expresses the default preference of responding with p_k over responding with p_j while the second rule states that under some specific conditions the preference is the other way round. More generally, we could have conditions NC_{kj} in the first rule that specify the normal conditions under which the default preference applies.

Using this level, it is then possible to discriminate between the dialogue moves by simply specifying that the agent will usually prefer his default behaviour, unless some special conditions are satisfied. Typically, the later situation can capture the fact the strategy should vary when exceptional conditions hold (for example when the others agents have specific roles). More generally this would cover any tactics pertaining to the roles of the agents *Y*, the content, *S*, and other relevant factors of the current situation.

Example 2. *Consider the following rules defining the tactic component of the agent Bob.*

$$
\begin{aligned}
R^{requ}_{acc|chall}(Y,S): &\quad h\_p(r_{acc,requ}(Y,S), r_{chall,requ}(Y,S)) \\
&\quad \leftarrow true \\
R^{requ}_{chall|acc}(Y,S): &\quad h\_p(r_{chall,requ}(Y,S), r_{acc,requ}(Y,S)) \\
&\quad \leftarrow unknown(Y,X) \\
R^{requ}_{acc|ref}(Y,S): &\quad h\_p(r_{acc,requ}(Y,S), r_{ref,requ}(Y,S)) \\
&\quad \leftarrow manager(Y,X) \\
R^{requ}_{chall|ref}(Y,S): &\quad h\_p(r_{chall,requ}(Y,S), r_{ref,requ}(Y,S)) \\
&\quad \leftarrow true
\end{aligned}
$$

Now assuming in the background knowledge $\mathcal{T}_0$ of Bob that Al is known to be a manager of Bob, then this tactical theory together with the basic component introduced in the previous example would give accept(Bob, Al, nail) as the only credulous and indeed skeptical consequence of the theory for the next reply move of Bob. The normal default preferences apply. However, if $\mathcal{T}_0$ of Bob contained unknown(Al, Bob) (and so Al was not a manager of Bob) then both accept(Bob, Al, nail) and challenge(Bob, Al, nail) would be credulously admissible and hence possible reply moves.
In order to overturn the default of accepting over challenging, in this specific context of unknown requesters, a rule at the third specific context level of the tactical theory would be needed. We would have in the set $\mathcal{P}_C$ of the tactical component the higher-level priority rule:

$$
C^{tactical}_{chall|acc}: \quad h\_p(R^{requ}_{chall|acc}, R^{requ}_{acc|chall}) \quad \leftarrow true
$$

Then the only possible move for Bob would be to challenge.

Note that the $T_{tactical}$ component of the personal strategy theory could change from application to application as the tactic that an agent may want to apply could be different. A designer may hold different tactic components and equip its agent with the relevant one, depending on the application. Alternatively, this flexibility could be captured in one theory $T_{tactical}$ by introducing suitable *tactical conditions* in these priority rules to separate the cases of different applications. For instance, the role of manager could be important in certain applications, but not in others. In this case the priority rule will be written as:

$$R^{requ}_{acc|ref}: \quad h\_p(r_{acc,requ}(Y,S), r_{ref,requ}(Y,S)) \leftarrow manager(Y,X), context(S)$$

where the $context(S)$ is the tactical condition that defines in $\mathcal{T}_0$ the situations (applications) where the management relation is significant.

The attitude component $T_{attitude}$. This third component of the private strategy theory of an agent captures general, typically application independent, charateristics of personal strategy that the agent applies. This consists of priority rules R and C (like the $T_{tactical}$ component) on the rules of the first component T_{basic}. They are again of the form:

$$R^{name}_{j|k}: \quad h\_p(r_{j,i}, r_{k,i}) \quad \leftarrow \quad b_{jk}$$

where i, j, k belong to the performative types label set, *Perf*. Here *name* is an identifier name for this personal strategy and b_{jk} are called *behaviour* conditions under which a particular personal strategy is defined. Higher-level C rules can be included on these R rules as above to allow the flexibility to deviate from a normal personal strategy under special circumstances.

Example 3. *Let us now consider the following attitute theory, we call* $T_{altruistic}$, *whereby agent Bob prefers to accept a request when it does not need the resource. This theory has the priority rule:*

$$R^{altruistic}_{acc|chall}: \quad h\_p(r_{acc,requ}, r_{chall,requ}) \quad \leftarrow \neg need(P,X)$$

Hence if the background theory is now extended to

$$\mathcal{T}_0 \cup \{\neg need(nail, Bob)\}$$

then Bob will give preference to the rule $r_{acc,requ}$, *and* $accept(Bob, Al, nail)$ *will be the skeptical conclusion.*

Conflicts between components. It is now important to note that the latter two components may have different priorities, that is the tactical component may give priority to a rule while the attitude component does the opposite. Consider for example an attitude theory, called $T_{argumentative}$, specifying the personal attitude that *Bob* prefers to challenge whenever possible as specified by [1] Amgoud and Parsons, 2001. We will examine later on in more details the link between our attitude components and the agent type strategies proposed in [1].

Example 4. $T_{argumentative}$ *would contain rules of the form:*

$$R^{argumentative}_{chall|acc}: \quad h\_p(r_{chall,requ}, r_{acc,requ}) \quad \leftarrow true$$

Then Bob under its personal attitude theory will always give preference to challenge. Hence both accept(Bob, Al, nail) and challenge(Bob, Al, nail) are credulous consequences of the overall strategy theory containing the tactic and attitude components.

Therefore, dilemmas (non-determinism) in the overall decision of our theory can exist. We can then use higher-level priority rules in the attitude component to resolve conflicts either way, in favour of *attitude dominance* or of *tactical dominance*. These special higher-order rules would then refer to R-rules in any of the components, i.e. also in $T_{tactical}$. In the case of our example, if we wanted to impose the attitude strategy we would then have a higher-order rule:

$$C^{argumentative}: \quad h\_p(R^{argumentative}, R^K) \leftarrow K \neq argumentative$$

Such a rule gives flatly priority of the attitute preference rules over those of the tactical component. This can be made more specific to apply only on some subset of rules. e.g. that refer to only some performatives. Also we again have the flexibility to make this dominance conditional on specific conditions pertaining to the current knowledge of the agent about its world, e.g. that the dominance of the argumentative attitude in our example arises only when there is a danger involved in the request.

3.1. *Properties of private strategies*

An agent upon receipt of a performative from a fellow agent will typically dispose of several options in order to reply. These options are obtained by computing (credulous or skeptical) conclusions of its strategy theory.

Often a desirable theoretical property of the strategy theory is that this is *non-concurrent*, namely that at most one dialogue move is generated at any time. In our framework, this is guaranteed by construction because every strategy includes rules making concurrent moves conflicting with each other. In other words, there is no admissible argument that would support two different moves. Observe that this property is often called *determinism* in similar frameworks [16], because the semantics used does not allow concurrent sets of admissible arguments. In our case, non-concurrency does not guarantee determinism in the usual sense. For instance, a *credulous* agent would typically pick up an admissible argument at random when facing different alternatives (and may then respond differently to the same performative).

To guarantee that at least one such admissible argument exists, we need to inspect the conditions that appear at the first level of the strategy. In other words, we need to check that the strategy is *exhaustive* in the sense that the conditions of at least one of its rules at level 1 are always satisfied. Again, this does not coincide exactly with the *existence* of a reply move [16]. For instance, a *skeptical* agent would not choose between different candidate moves (admissible arguments), and remain silent (if there are no moves generated then we can have a special utterance $\mathcal{U}$, indicating that this is the case [11] and either the dialogue would terminate or suspend until more information is acquired by the agent).

One way to ensure that all these notions actually coincide is to require that the complete strategy theory, comprising of all its three components together, has a *hierarchical* form defined as follows.

Definition 3.1. (Hierarchical Policy). *An agent's argumentative policy theory, S, is hierarchical iff for every pair of rules* s_i, s_j *in S whose conclusions are incompatible, there exists a priority rule,* p_i^j *in S, that assigns higher priority to one of these two rules, such that, whenever both the conditions of* s_i, s_j *are satisfiable (in the background theory of S) so is the condition of* p_i^j.

Note that in this definition the rules s_i, s_j could be themselves priority rules in which case the rule p_i^j is a priority rule at a higher level. Basically, the hierarchical structure prevents the existence of concurrent sets of admissible arguments. In this case, of course, the (unique) credulous conclusion and the skeptical conclusion would coincide. As a consequence, non-concurrency implies determinism, and exhaustivness implies existence. This leads directly to the following result:

Theorem 3.1 (Uniqueness). *If the strategy theory is exhaustive, hierarchical and its priority relation does not contain any cycles of length* > 2, *then the agent will always have exactly one move to utter in its reply.*

4. Flexible society protocols

We now turn to the representation of society protocols. Protocols specify what is deemed legal for a given interaction, that is which dialogue moves can follow up after a (sequence of) dialogue move(s). We shall see how protocols can be specified using the same logical framework as for the private strategies in an analogous way, as argumentation theories divided in three parts. Note that there is no issue of determinism here. A protocol will typically allow an arbitrary number of legal continuations: any credulous consequence of the society protocol theory would be a legal move. However, exploiting the flexibility of our framework to take into account exceptional situations that may arise in interactions, we shall introduce different notions of legality.

In the first part (P_0), we specify all the dialogue moves that may be legal in some circumstances, namely the possible legal follow-ups after a dialogue move $p_i(Y, X, S)$. By defining one such a rule

$$r_{j,i}(Y, S) : p_j(X, Y, S') \leftarrow p_i(Y, X, S), S_{ij}$$

for each possible legal continuation under the conditions S_{ij} which in the simplest case can be taken to be empty. Note that this lower-level part of the protocol is completely analogous to the basic component of the private strategy theory of an agent and in some cases it can be replaced by it. At this level then we have several single moves as credulous conclusions and hence legal moves. We will refer to this set as the set of *potentially legal* moves.

Example 5. *Consider for instance the following protocol which regulates requesting interactions (observe that this protocol does not cater for counter-proposals).*

$$\begin{array}{llcl} r_{acc,req}: & accept(X,Y,P) & \leftarrow & request(Y,X,P) \\ r_{ref,req}: & refuse(X,Y,P) & \leftarrow & request(Y,X,P) \\ r_{chal,req}: & challenge(X,Y,P) & \leftarrow & request(Y,X,P) \end{array}$$

The set of potentially legal moves clearly contains accept, refuse and challenge.

The main task of the protocol is then to specify which of the potentially legal moves are in fact legal under normal circumstances. This is done by representing a preference policy at the next part (P_1) of the society protocol theory, whose rules have the form:

$$R^{l}_{j|k} : h\_p(r_{ji}, r_{ki}) \leftarrow N_{jk}, \quad l = 1, 2 \ldots$$

where N_{jk} are conditions that hold in a normal situation. Such a rule gives priority of the move p_j over p_k under the conditions N_{jk} and hence in the absence of any other rule it will render p_k illegal, as this is not a credulous conclusion of the full ($P_0 \cup P_1$) protocol theory now. Note that unlike conditions appearing in the agents' strategies, these protocol conditions are assumed to be objective and verifiable. We will assume that these conditions should hold in the (shared) commitment store (CS) of the agents involved in the interaction.

We can then define the set of *normal (or default) legal* moves as those moves that are credulous consequences of the theory $CS \cup P_0 \cup P_1$.

Example 6. *The (normal) preference policy rules regulating the delivering of a drug are the following: (i) if the prescription is shown then you can accept to give the drug, (ii) if the request is from a child then refuse to provide the drug, and (iii) in any case you are allowed to challenge the request. This protocol can be captured by the rules* ($k \in Perf$):

$$\begin{array}{ll} R^{1}_{accept|k}: & h\_p(r_{accept,request}(Y,P), r_{k,request}(Y,P)) \\ & \leftarrow prescription(Y,P), k \neq accept \\ R^{2}_{refuse|k}: & h\_p(r_{refuse,request}(Y,P), r_{k,request}(Y,P)) \\ & \leftarrow child(Y), k \neq refuse \\ R^{3}_{challenge|k}: & h\_p(r_{challenge,request}(Y,P), r_{k,request}(Y,P)) \\ & \leftarrow k \neq challenge \end{array}$$

Let us now consider different cases: if $prescription(Al, drug)$ *holds in* CS*, then both* $accept(Bob, Al, drug)$ *and* $challenge(Bob, Al, drug)$ *are credulous conclusions. If* $child(Al)$ *holds in* CS *then both* $refuse(Bob, Al, drug)$ *and* $challenge(Bob, Al, drug)$ *are credulous conclusions. Finally, if it is the case that both* $prescription(Al, drug)$ *and* $child(Al)$ *holds in* CS *then all the potentially legal moves are again credulous conclusions. These are the normal or default legal moves under the aforementioned respective normal circumstances.*

In some particular situations we may want the protocol to impose a special requirement that could render some normally legal moves illegal, or even some illegal moves legal. To have this added flexibility we can complete our protocol theory with a third part (P_3) that contains priority rules that apply under special situations. Some of these are higher-order priority rules on the other priority rules. The rules of P_3 will have the form:

$$C_{k|j} : h\_p(R^m_{k|j}, R^n_{j|k}) \leftarrow E^C_{kj}, \quad m, n = 1, 2 \ldots$$
$$R_{k|j} : h\_p(r_{ki}, r_{ji}) \leftarrow E^R_{kj}$$

where E^R_{kj} are conditions describing *special conditions* and similarly E^C_{kj} are *special situations* that give priority of $R^m_{k|j}$ over $R^n_{j|k}$.

We are now in position to define the set of *exceptional legal* moves as those moves that are credulous consequences of the theory obtained by conjoining CS together with the overall society component ($P_0 \cup P_1 \cup P_2$).

Example 7. *The protocol is now refined by requiring that if the drug is toxic then a child's request should be refused. This is captured by*

$$C^{toxic}_{ref|chall} : \quad h\_p(R^2_{ref|chall}(Y, P), R^3_{chall|ref}(Y, P)) \leftarrow toxic(P)$$

Then in the full protocol theory the move challenge(Bob, Al, drug) when Al is a child is not a credulous consequence any more and the only exceptional legal move is then refuse(Bob, Al, drug).

Observe that it is possible to have moves that are normally illegal become exceptionally legal as illustrated by the following example.

Example 8. *The protocol is now refined by specifying that (i) if the request is urgent then it should be allowed to accept it, and (ii) if it is also critical then the seller must accept the request.*

$$R^{urgent}_{acc|k} : \quad h\_p(r_{acc,requ}(Y, P), r_{k,requ}(Y, P)) \leftarrow urgent(P), k \neq accept$$
$$C^{critical}_{acc|j} : \quad h\_p(R^{urgent}_{acc|j}(Y, P), R^m_{j|acc}(Y, P)) \leftarrow critical(P)$$

With this added to the protocol theory, the move accept(Bob, Al, drug) when Al is a child becomes a credulous consequence if urgent(P) holds in CS. If critical(P) also holds, then it is even a skeptical conclusion.

In our framework, the reference to conditions allows us to define the circumstances under which the potentially legal moves are normally or exceptionally legal. Interestingly, the status of legality is non monotonic under new information on these conditions. As the information kept in the commitment stores will evolve during the dialogue, it can even become a matter of discussion for the agents.

5. Related work and conclusions

There is an increasing amount of work on argument-based interaction, mainly focused on negotiation—see [15] for a survey. More generally, according to [11], apart from its naturalness, an argumentation-based approach has two major advantages: rationality of the agents, and a social semantics in the sense of [17]. Our argumentation-based approach inherits these advantages in adressesing both the private aspects of agents' strategies, along with the social aspects of interaction protocol and providing added flexibility. Agent strategies give adaptable behaviour according to the context of the dialogue and the particular roles of the participating interlocutors. At the social level, flexible protocols can be defined that can cater for a wide variety of interactions, including specific circumstances that may come up as the dialogue evolves.

5.1. Agents' profiles

Different notions of agent profiles have been proposed in the literature. Amgoud and Parsons [1], for instance, have proposed five profiles of dialogues to discriminate between different classes of agent types with varying degree of "willingness to cooperate" in the personal attitude of an agent. These profiles are: *agreeable* (accept whenever possible), *disagreeable* (only accept when there is no reason not to), *open minded* (only challenge when necessary), *argumentative* (challenge whenever possible), *elephant child* (question whenever possible). The enhanced flexibility of our approach allows us to capture these profiles as special cases.

Theorem 5.1. *The agent type strategies of* [1]: "agreeable", "disagreeable", "argumentative", "open-minded", and "elephant child", can be captured as private agent strategies.

Proof. We start by considering the *agreeable* attitude (accept whenever possible). We can capture this as folllows: whenever the (or a) dialogue step leading to accept is enabled (so its rationality conditions are satisfied) then this would have higher priority than other dialogue steps. This is easily expressed by the following rules in the second level of the attitude component of a strategy:

$$R^{agreeable}_{accept|k}: \quad h\_p(r_{accept,i}, r_{k,i}) \quad \leftarrow \quad k \neq accept$$

for every $i, k \in Perf$. This gives the agreeable strategy in the cases when the second component $T_{tactical}$ of the private strategy theory is empty. Otherwise, we could have rules, $R^{tactical}_{k|accept}$, which could make the move also possible. To impose the agreable strategy we include in the attitude component the higher-order rule

$$C^{agreeable}: \quad h\_p(R^{agreeable}_{accept|k}, R^{tactical}_{k|accept}) \quad \leftarrow \quad k \neq accept$$

for every rule $R^{tactical}_{k|accept}$ of the tactical component.

Similarly, we can capture the other agent type strategies. First, we can easily observe that *Argumentative/Elephant's Child* attitudes are completely analogous to *Agreeable* by replacing *accept* by *challenge* and *question,* respectively. As for the *Disagreeable* attitude, we must interpret this attitude as folllows: for every i and k in the label set *Perf*, any dialogue step $r_{k,i}$ that is enabled will be preferred over $r_{accept,i}$ to follow the current performative p_i.

$$R^{disagreeable}_{k|accept}: \quad h\_p(r_{k,i}, r_{accept,i}) \quad \leftarrow \quad k \neq accept$$

for every i and k in the label set *Perf*, together with the analogous $C^{disagreeable}$ higher-order rule. So only when no $r_{k,i}(k \neq accept)$ rule is enabled can *accept* be the next move as otherwise this other k-move will be preferred, as any condition that enables $r_{k,i}(k \neq accept)$ is a reason not to accept.

We conclude with the *Open-minded* attitude. How to capture this attitude depends on what interpretation we give to "when necessary". It could alternatively be "when there is no other move possible" or "when the *accept* move is not enabled as a next possible move". In any case, we can choose one and capture it as shown above. □

We also conjecture that it would be possible to formalize the different "assertion and "acceptance" attitudes and consequently the different agent profiles (i.e. confident, careful thoughtful and credulous, cautious, skeptical, respectively) proposed in the recent work of [11].

5.2. *Cognitive agent architectures*

Another related work is that of the *BOID architecture* [3]. This defines several agents types (e.g. realistic, selfish, social, etc.) depending on the priority the agent gives to these different *mental attitudes* (Beliefs, Obligations, Intentions, Desires). This is related to our approach whereby the agent can solve conflicts between components of its theory. The society protocols can be considered as the normative aspect of the system, whereas the tactical component is more related to intentions and desires. Different meta-level preferences of these components would give agents of different types. Note that our framework allows for argumentation to be carried out also on the conditions in agents' strategies. These can then be considered as part of the agent's beliefs and hence our agents are *realistic* in the sense of [3].

5.3. *Logic-based protocols*

In Endriss et al. [4], protocols are translated into integrity constraint rules, in Abductive Logic Programming (ALP), of the form $p_i \Rightarrow \vee p_j$. These can easily be translated into rules at the first level of our protocols. It is instructive though to ask the reverse question of how would this ALP-based approach capture our theories. Both representation formalisms exhibit a closely related expressiveness. Clearly, it would be possible to cater for exceptional situations by introducting "abnormality" predicates in ALP rules. But the two approaches use different logical notions for the semantics of the protocol: logical consistency for the ALP-based and (non-deterministic) admissibility for our argumentation-based approach. The non-locality of the consistency requirement (any single conflict in the integrity constraints would render the whole protocol theory inconsistent and all moves illegal) suggests that this solution would be highly non-modular. The drawback of this approach is indeed that it would be necessary to modify a large amount of rules each time an exceptional situation is identified by the designer. The approach described in this paper, on the other hand, allows for a completely modular specification. That is, a designer using our framework could make use of general libraries of strategies, but still specify how the specific context should affect these rules (without having to modify them though).

5.4. Commitment machines

In Yolum and Singh [18], *social commiments* are used as a way to specify protocols by refering to the content of the actions. By allowing reference to the content of the moves (and other relevant information contained in the commitment store), we cater for the kind of flexibility discussed in [18]. However, our approach is closer in spirit to *dialogue games* approaches where dialogue rules and conditions on the commitment stores are used in combination to define the notion of legality. Further work is needed to evaluate how our approach compares to these commitment-based approaches.

In conclusion, our approach provides a way of realizing together several notions of argumentation-based communication that combines the merits of (a) modular separation of concerns, (b) great expressivity of the theories and (c) feasible implementation directly from their declarative specification. Further work is needed to develop a more systematic methodology for building these theories, for instance the design issue of how criteria should be distributed amongst the three components of the framework. Preliminary rules of thumb can be given (e.g. the attitude component relates to the domain independent personality of the agent that captures generic strategies of decision), but a more comprehensive account needs to be worked out.

Acknowledgments

We would like to thank Ulle Endriss and the anonymous referees of this paper for their comments. This work was partially supported by the IST programme of the EC, FET under the IST- 2001-32530 SOCS project, within the Global Computing proactive initiative.

References

1. L. Amgoud, and S. Parsons, "Agent dialogue with conflicting preferences," In: J.-J. C. Meyer and M. Tambe (eds.), *Intelligent Agent VIII. Revised papers of the International Workshop on Agent Theories, Architectures and Languages (ATAL-2001)*, Vol. 2333 of *Lecture Notes in Artificial Intelligence (LNAI)*, Springer-Verlag, pp. 190–205, 2001.
2. G. Brewka, "Dynamic argument systems: a formal model of argumentation based on situation calculus". *J. Logic Comput*. Vol. 11, no. 2, pp. 257–282, 2001.

3. J. Broersen, M. Dastani, J. Hulstijn, Z. Huang, and L. van der Torre, "The BOID architecture: conflicts between beliefs, obligations, intentions and desires". In: J. P. Mller, E. Andre, and S. Sen (eds.), *Proceedings of the 5th International Conference on Autonomous Agents (AGENTS-2001)*, pp. 9–16, 2001.
4. U. Endriss, N. Maudet, F. Sadri, and F. Toni, "Protocol conformance for logic-based agents," in: G. Gottlob and T. Walsh (eds.), *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-2003)*, Morgan Kaufmann Publishers, pp. 679–684, 2003.
5. Foundation for Intelligent Physical Agents (FIPA), "Communicative Act Library Specification (XC00037H)," 2001, http://www.fipa.org/spec.
6. GORGIAS, A system for argumentation and abduction, 2002, 'http://www.cs.ucy.ac.cy/ nkd/gorgias'.
7. C. L. Hamblin, "Fallacies", 1970, Methuen.
8. A. Kakas, P. Mancarella, and P. M. Dung, "The acceptability semantics for logic programs" In: P. V. Hentenryck (ed.), *Proceedings of the 11th International Conference on Logic Programming*, pp. 504–519, MIT Press, 1994.
9. A. Kakas and P. Moraitis, "Argumentation-based decision making for autonomous agents". in: J. S. Rosenschein, T. Sandholm, M. Wooldridge, and M. Yokoo (eds.), *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2003)*, ACM Press, pp. 883–890, 2003.
10. N. Karacapilidis and P. Moraitis, "Engineering issues in interagent dialogues," in: F. V. Harmelen (ed.), *Proceedings of the 15th European Conference on Artificial Intelligence (ECAI-2002)*, IOS Press, pp. 58–62, 2002.
11. P. McBurney and S. Parsons, "Argumentation-Based Communication between Agents". In: M.-P. Huget (ed.), *Communication in Multiagent Systems*, Vol. 2650 of *Lecture Notes in Artificial Intelligence (LNAI)*, Springer-Verlag, 2003.
12. P. McBurney, S. Parsons, and M. Wooldridge, "Desiderata for agent argumentation protocols". In: C. Castelfranchi and W. L. Johnson (eds.), *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2002)*, ACM Press, pp. 402–409, 2002.
13. S. Parsons, C. Sierra, and N. R. Jennings, "Agents that reason and negotiate by arguing". *J. Logic Comput.,* Vol. 8, no. 3, pp. 261–292, 1998.
14. H. Prakken and G. Sartor, "A dialectical model of assessing conflicting arguments in legal reasoning". *Artificial Intelligence Law,* Vol. 4, pp. 331–368, 1996.
15. I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and L. Sonenberg, "Argument-based negotiation," *Knowledge Eng. Rev*. Vol. 18, no. 4, pp. 343–375, 2003.
16. F. Sadri, F. Toni, and P. Torroni, "Dialogues for negotiation: agent varieties and dialogue sequences," In: J.-J. C. Meyer and M. Tambe (eds.), *Intelligent Agent VIII. Revised papers of the International Workshop on Agent Theories, Architectures and Languages (ATAL-2001)*, Vol. 2333 of *Lecture Notes in Artificial Intelligence (LNAI)*, Springer-Verlag, 2001.
17. M. P. Singh, "A social semantics for agent communication language," In: F. Dignum and M. Greaves (eds.), *Issues in Agent Communication*, No. 1916 in *Lecture Notes in Computer Science (LNCS)*, Springer-Verlag, pp. 31–45, 2000.
18. P. Yolum, and M. P. Singh, "Flexible protocol specification and execution: applying event calculus planning using commitments," in: C. Castelfranchi and W. L. Johnson (eds.), *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2002),* ACM Press, pp. 527–534, 2002.