Argumentation based Modelling of Embedded Agent Dialogues

Yannis Dimopoulos, Antonis C. Kakas, and Pavlos Moraitis

Department of Computer Science, University of Cyprus
P.O. Box 20537, CY1678, Nicosia, Cyprus
yannis,antonis,moraitis@cs.ucy.ac.cy

Abstract. This paper presents a novel approach to modelling embedded agent dialogues. It proposes a specific structure for the supporting information accompanying the arguments that agents exchange during a dialogue, it defines formally how this information relates to the agent theory, and assigns to it semantics that is associated to each of the atomic dialogue types of the Walton-Krabbe typology. This allows the formal definition of necessary and sufficient initiation and acceptance conditions of licit dialectical shifts that are necessary for the modelling of embedded agent dialogues.

1 Introduction

The task of modelling agent dialogues has proved to be of great importance in representing complex agent interactions. Since the work of Walton and Krabbe [16] proposing a classification of possible atomic dialogue types (i.e. deliberation, negotiation, persuasion, information-inquiry, information-seeking, eristic) a lot of work have been devoted to modelling the first five of them, the sixth being considered unappropriate in a multi-agent context. Recently, some of this work has adopted an argumentation-based approach for such dialogue modelling as can be found for example in [13], [1], [14], [6], [11], [12]. However, to our knowledge, there exists only a few cases (see e.g. [10], [8], [14]) of study of the combination of atomic dialogues and of the particular combination of embedded dialogues.

Embedded dialogues are a very interesting combination of atomic dialogues. They concern situations where during a specific dialogue type, the interlocutors can shift to another dialogue type. When this subsidiary dialogue closes a shift back is made to the external dialogue which will continue from the point where it was interrupted. As Walton [15] says: "the one dialogue can be "sandwiched in" between the prior and subsequent parts of an enveloping sequence of dialogue of another type. Practical reasons can cause the interruption, but then the dialogue can quickly shift back to the original type". In the case of embedded dialogues the outcome of the second dialogue can influence the quality of the outcome of the original dialogue, because the second dialogue is functionally related to the argumentation in the first dialogue.

An important issue in the multi-agent context, is related to the ability of detecting in a current dialogue, licit dialectical shifts, which according to the

literature (see e.g. [15]), are those that allow agents to transit to another type of dialogue which supports the old goals or at least allows their fulfilment to be carried forward. Such dialogues shifts to embedded dialogues are useful in contributing to the successful completion of the outer dialogues. If the new dialogue is blocking the old goals, the dialectical shift is considered illicit and it is often associated with informal fallacies [15] which we believe are less appropriate for artificial agents dialogues.

In this paper we investigate how to model embedded dialogues based on the argumentation reasoning of the agent. We present an argumentation framework in which an agent represents and reasons with the various components of its knowledge and dialogue theory. Based on this the agent is equipped with a set of different capabilities for reasoning about goals, beliefs and actions. We then define formally the structure of the supporting information accompanying the exchanged arguments between the agents during a dialogue and present how its constituents are related to goals, beliefs and actions. This allows us to link the argumentation-based reasoning of the agent to its dialogues and formalize within the framework the five atomic dialogue types of the Walton-Krabbe typology. In turn we can give a formal notion of licit dialectical shifts (in the context of embedded dialogues) through the definition of initiation conditions for the five atomic dialogue types and acceptance conditions for such dialectical shifts. To our knowledge, our work is one of the first attempts to provide formal definitions for all these issues related to the modelling of embedded dialogues.

The rest of the paper is organized as follows. Section 2 presents briefly the underlying argumentation theory and the primitive components that a framework needs to possess in order to build embedded dialogues. Section 3 defines the dialogue supporting information while section 4 presents the embedded dialogue framework we propose. Finally, section 5 discusses related work and concludes.

2 Background

2.1 Basic Argumentation Theory

This section gives briefly the basic concepts of the underlying argumentation framework in which an agent represents and reasons with its communication theory. With this the agent will be able to generate, and then communicate to other agents, different arguments for the various topics involved in its dialogues. There are two important features of an argumentation framework that are required for this purpose. Firstly, the framework needs to be *adaptive* to changes in the current knowledge of the agent about the state of the world. Secondly, the framework should be able to identify in its arguments a set of (significant) conditions on which an argument is *supported*. In particular, this set may contain assumptions pertaining to the incomplete information that the agent has about the world. Any argumentation framework that can provide these two functions is suitable.

An argumentation framework in its abstract form is based on a set, $\mathcal{A}$, of arguments and a binary attacking relation, $\mathcal{AR}$, amongst these arguments. We

will assume that arguments in $\mathcal{A}$ are represented by logical theories in some background monotonic logic whose derivability relation we will denote by $\vdash_B$. Each argument A is a subset of a given theory $\mathcal{T}$ and we say that A is an *argument for* L whenever $A \vdash_B L$. An example of such a framework, called Logic Programming without Negation as Failure ($LPwNF$), was proposed in [5] in which theories are written in terms of Extended Logic Programming rules and priorities on these rules. This was developed further in [7] providing the above desired features of adaptability and supportedness of arguments. Although not crucial for the work in this paper we will adopt this framework in order to be more concrete in our presentation.

In this framework of $LPwNF$ the attacking relation $\mathcal{AR}$ is realized via a (symmetric) notion of *incompatibility* between literals, that defines when two literals cannot hold together, and a set of priority rules, given within the same theory $\mathcal{T}$. Informally, given two subsets A', A of $\mathcal{T}$, A' attacks A if they have incompatible consequences under the background logic $\vdash_B$ and A' is stronger than A according to the priority rules in the theory. Thus a given argumentation theory $\mathcal{T}$ defines both the set of arguments and the attacking relation amongst them.

The central notion for the acceptance of an argument is that of *admissibility*. This and the argumentation entailments that follow from it are defined as follows.

Definition 1. *Let $\mathcal{T}$ be an argumentation theory. An argument $\Delta \subseteq \mathcal{T}$ is* admissible *iff Δ does not attack itself (it is consistent) and for any $\Delta' \subseteq \mathcal{T}$ if Δ' attacks Δ then Δ attacks Δ'.*
Given a literal L then L is a skeptical consequence *of the theory iff L holds, under the background monotonic logic $\vdash_B$, in an admissible subset of $\mathcal{T}$ and for any literal, $\overline{L}$, which is incompatible with L, there exists no admissible argument in which $\overline{L}$ holds under $\vdash_B$.*

In several cases we want to base the admissibility of an argument on some significant information about the specific case in which we are reasoning or on incomplete information that is missing from our theory. We can formalize this conditional form of argumentative reasoning by defining the notion of *supporting information* and extending argumentation with abduction on this information.

Definition 2. *Let $\mathcal{T}$ be an argumentation theory and, Ab, a distinguished set of predicates in the language of $\mathcal{T}$, called* abducible *predicates. Given a literal L, a* supported argument *for L is a tuple (Δ, S), where S is a set of ground abducible facts not in Δ such that Δ is not an admissible argument for L, but $\Delta \cup S$ is an admissible argument for L. We say that S is supporting information for the argument Δ of L.*

Given this we have an argumentation entailment, $\vdash_{arg}$, defined as follows.

Definition 3. *Let $\mathcal{T}$ be an argumentation theory and, Ab, a distinguished set of* abducible *predicates. Given a literal L, $T \vdash_{arg} L$, iff there exists a set of ground abducible facts S such that L is a skeptical consequence of $\mathcal{T} \cup S$. In other words, there exists in $\mathcal{T}$ a supported argument (Δ, S) for L and for any literal $\overline{L}$ which is incompatible with L there exists no supported argument for $\overline{L}$ in $\mathcal{T} \cup S$.*

2.2 Primitives for Embedded Dialogues

In this section we present the primitives components that a framework needs to possess in order to build embedded dialogues within this framework.

Reasoning capabilities of an agent Dialogues depend on the reasoning capabilities of the agents. We consider that different reasoning capabilities are involved in the reasoning process of the agents, during the different possible types of dialogues. This reasoning process may concern a goal decision reasoning capability for the choice of the preferred goal to be achieved, a temporal reasoning capability about actions and change for deriving its beliefs about the current (or future state of the work) and a plan preference capability for deriving preferred plans for a goal. In this paper we consider that all these different capabilities can be derived via suitable argumentation theories in the argumentation framework described above. The importance of using argumentation as a basis for the reasoning capabilities stems from the fact that agents can then exchange, during their dialogue, their arguments (and the supporting information for these) and use these to develop their dialogues.

We will assume that agents have the following argumentation based capabilities that operate on their knowledge T:

- a preferred plan capability, $\vdash_{PPlan}$, which is given by the synthesis of planning capability; $\vdash_{Plan}$, and a plan preference selection,$\vdash_{PP}$. Hence given a goal G, if $T \vdash_{PPlan} plan(G)$ then $plan(G)$ is a preferred plan for the goal G. We will also write $T \vdash_{PPlan} G$ to mean that there exists a preferred plan for the goal G.
- a desired goal capability, $\vdash$, that derives goals which are currently preferred by the agent and for which it also has a (preferred) plan to satisfy. Hence, $T \vdash G$ can be decomposed into a *goal decision* capability, $T \vdash_{GP} G$, and the $T \vdash_{PPlan} G$.
- a temporal reasoning or Reasoning about Actions and Change (RAC) capability, $\vdash_{RAC}$, with which the agent is able to derive its beliefs about the current (or future) state of the world. This is based on the agent's knowledge of action effect laws (and constraints) and on narrative knowledge about the past containing actions that have occurred and past observations of properties of the world.

Dialogue supporting information The supporting information accompanying the arguments the agents exchange during a dialogue is important for various reasons. For example, we will see that it helps characterize the type of the appropriate dialogue to be undertaken according to the topic to be discussed at a certain instant of a specific dialogue type. In this paper we will structure the supporting information according to how it relates to the goals, beliefs and actions of the agents. Supporting information comes from the underlying argumentation reasoning (as described above) that is used to implement the reasoning capabilities of the agent.

For the planning capability $\vdash_{PPlan}$ any generated plan, $plan(G)$, itself forms supporting information in the form of future actions for the arguments that derive the goal G. (Note that part of the plan can be requests for other agents to achieve a needed subgoal for G.) For the desired goal capability $\vdash$ an admissible argument will contain in its support conditions for the goal to be both desired and have a preferred plan (intention) under which the agent aims to achieve it. Finally, for the temporal reasoning capability the support, S, of arguments for current beliefs contains assumptions on properties at earlier times which are unknown to the agent and therefore it needs to hypothesize these.

We will see below that (part of) this support will be communicated with the aim to inform the other agent the *Reasons* why the agent wishes to achieve the goal G in this way and the *Terms* that the agent requires from the other agent in its endeavor for G.

Atomic dialogues initiation conditions A specific dialogue type can be initiated only under certain necessary conditions. We will propose a formal definition of such initiation conditions for the five atomic dialogues types of the Walton-Krabbe typology, based on a synthesis of informal descriptions proposed in the literature, but we do not pretend that these conditions may cover the totality of the possible situations. The definition of these initiation conditions will be based on arguments of the agents for their desired goals and their supporting information.

Dialectical shift A dialectical shift (see e.g [15]) is a transit from a certain type of dialogue to another of different type. This transit might allow agents to achieve goals whose fulfillment was impossible in the originally open dialogue. The definition of such a dialectical shift corresponds to a set of sufficient conditions under which such a transit is possible. The acceptance conditions of such a shift must also be defined. Our definitions for these will again be based on the arguments and supporting information exchanged during the dialogue so far. Here again, we will not claim that our formalization is complete but rather that it forms a core that can be extended as needed for increasingly complex situations.

3 Dialogue Supporting Information

Agents operate in a dynamic and ever changing world. To keep track of the change an agent uses his capability, $\vdash_{RAC}$, to derive conclusions about how the world is in its current state (to simplify our discussion we assume that an agent never needs to reason about the past). We call *current atoms* (literals) the atoms (literals) of the theory of the agent that refer to the current state of the world. A current literal p is called a *belief* if $T \vdash_{RAC} p$

An agent can execute *actions* that can change the current state of the world to some other more "desirable" state. This new state is described via the set of *goals* that the agent wishes to achieve through the execution of actions. Atoms (literals) that refer to some future state of the world are called *future atoms*

(literals). A goal G is a conjunction or set of future literals some of which are not true in the current state of the world, such that $T \vdash G$.

We will call *locutions* or *dialogue moves* the sentences that are exchanged between the agents during a dialogue. Locutions are 4-tuples of the form $P(a, b, t, Content)$ where P is a *performative* contained in a set that is in the lines of those used in [2], a is the agent that utters the locution, b is the intended recipient of the locution and t specifies the type of the current dialogue $\mathcal{D}$ the locution is uttered or the type of the dialogue to be initiated by the current locution. The $Content$ of the message is a 3-tuple of the form $< topic, reason, terms >$ where *topic* concerns the subject of the specific dialogue and it may be a set of goals, beliefs or actions of the involved agents and the other, possibly empty, fields correspond to the *supporting information* of the argument proving the literals contained in the field topic.

The proposed structure for the supporting information is partially inspired by the work presented in [11]. The literals that appear in the set *reason* refer to what the agent believes is true in the current state of the world, whereas the literals in the *terms* refer to what must be true in the future so that his goals succeed. More specifically, the set *terms* is the union of two subsets TR^- and TR^+ with the following meaning. If $p \in TR^-$, then for any other agent β it must be the case that $T_\beta \nvdash \neg p$ whereas if $p \in TR^+$ then for some other agent β it must be the case that $T_\beta \vdash p$. Intuitively, the literals in TR^- refer to actions or goals that the other agents should refrain from executing or achieving, whereas the literals in TR^+ refer to actions or future literals that the agent requests that other agents will execute or achieve.

In a similar way, the set *reasons* of an agent α is divided in two subsets, R^K and R^U. The set R^K contains a current literal p iff $T_\alpha \vdash_{RAC} p$, whereas R^U contains current literals that are assumptions made by agent α. By placing a current literal p in the set R^U of a locution, an agent declares that he assumes that p has the value true as he has no sufficient information from which he can derive the value of p. Therefore, the content of a locution is a 3-tuple of the form $< TP, < R^K, R^U >, < TR^+, TR^- >>$, where TP is the topic as noted above.

In this paper we assume that the agents are truthful, in the sense that the information they communicate with other agents is a consequence of their knowledge bases. Formally, if $P(a, b, t, < TP, R, TR >)$ is a locution sent by agent a to agent b it must be the case that the theory T_a of agent a has an admissible argument (Δ_a, S_a) such that $(\Delta_a, S_a) \vdash TP$ and $R \cup TR \subseteq S_a$.

4 The embedded dialogue framework

In this section we present formally our framework for embedded dialogues. We will restrict our attention to dialogues between two agents. In this context a dialogue is defined as follows.

Definition 4. *Dialogue*
A dialogue $\mathcal{D}$ between agents α and β is a finite sequence of the form $\mathcal{D}$=$L_1^{k|l} L_2^{l|k} \ldots$

$L_m^{j|n}$ *with* $k, l \in \{\alpha, \beta\}$*, where each element* $L_i^{p|q}$*, called the* i dialogue step*, is a locution of the form* $P(p, q, t, C)$*, and* $j = k$*,* $n = l$ *if* m *is odd and* $j = l$*,* $n = k$ *if* m *is even.*

We define now the outcome of a dialogue, and its sub-dialogues, for each of the participating agents. This definition is in line with the one presented in [13].

Definition 5. *Dialogue Outcome*
Let $\mathcal{D}$=$L_1^{k|l} L_2^{l|k} \ldots L_m^{j|n}$ *be a dialogue between agents* α *and* β*, with* $L_i^{p|q} = P^i(p, q, t, < TP_i, < R_i^K, R_i^U >, < TR_i^+, TR_i^- >>)$*. The outcome of* $\mathcal{D}$ *for agent* α *is defined as the set* $O_{\mathcal{D}}^{\alpha} = \bigcup_{i=1}^{m} \{s | s \in TP_i \cup R_i^K \cup TR_i^-$*, for* $L_i^{\alpha|\beta} \in \mathcal{D}$ *and* $P^i = accept\}$*. Similarly, the outcome of* $\mathcal{D}$ *for agent* β *is the set* $O_{\mathcal{D}}^{\beta} = \bigcup_{i=1}^{m} \{s | s \in TP_i \cup R_i^K \cup TR_i^-$*, for* $L_i^{\beta|\alpha} \in \mathcal{D}$ *and* $P^i = accept\}$*.*
Given a dialogue $\mathcal{D}$=$L_1^{k|l} L_2^{l|k} \ldots L_r^{c|d} \ldots L_m^{j|n}$ *between agents* α *and* β*,* $O_{\mathcal{D}_r}^{\alpha}$ *denotes the outcome for agent* α *of the sub-dialogue that starts at step 1 and ends at step* r*, and is defined as* $O_{\mathcal{D}_r}^{\alpha} = O_{\mathcal{D}'}^{\alpha}$*, where* $\mathcal{D}'$ *is the dialogue* $\mathcal{D}'$=$L_1^{k|l} L_2^{l|k} \ldots L_r^{c|d}$*. The definition of* $O_{\mathcal{D}_r}^{\beta}$ *is similar.*

The theory of an agent, and therefore his beliefs, goals and plans, change during the course of a dialogue. These changes are realized via the function $rev(T, S)$ that takes a theory T and a set of literals S and revises T to a new theory T' so that $T' \vdash s$ for all $s \in S$.

Definition 6. *Agent Theories and Agent Goals*
If $\mathcal{D}$ *is a dialogue between agents* α *and* β*,* $T_{\alpha}^{\mathcal{D}_i}$ *denotes the theory of agent* α *at step* i *of the dialogue* $\mathcal{D}$*, and is defined as* $T_{\alpha}^{\mathcal{D}_i} = rev(T_{\alpha}, O_{\mathcal{D}_i}^{\alpha})$*, where* T_{α} *is the theory of agent* α *at the beginning of the dialogue.*
The goal of agent α *at step* i *of dialogue* $\mathcal{D}$ *is denoted by* G_{α}^{i} *and is a set of future literals such that* $T_{\alpha}^{\mathcal{D}_i} \vdash G_{\alpha}^{i}$*.*

4.1 Modelling dialectical shifts

In this subsection we present formal definitions for the initiation conditions of the five dialogue types of the Walton-Krabbe typology, the notion of licit dialectical shift [15], the acceptance conditions of such a shift and the notion of efficient dialectical shift. These definitions aim to capture informal descriptions, commonly accepted in the literature. The initiation conditions allow an agent to detect the possibility of a shift from the current dialogue to another dialogue of a different type. They are necessary conditions for a dialogue shift to occur. The initiation conditions are linked to the constituents of the content of the locutions exchanged between agents, which correspond to the supporting information of the arguments used by the agents during a dialogue.

A dialectical shift from a dialogue of any type different than negotiation to a negotiation dialogue means that either the participating agents have conflicting goals (or interests) (see e.g. [11]) or the terms in the locution of one of the agents

leads to the failure of the goals of the other agent. This is a more general consideration for negotiation than the one proposed in the Walton and Krabbe typology where negotiation concerns the division of some scarce resource. Formally, this type of shift is defined as follows.

Definition 7. *Negotiation*
Let α and β be two agents involved in a dialogue $\mathcal{D} = L_1^{k|l} L_2^{l|k} \dots L_i^{\beta|\alpha}$, with $L_i^{\beta|\alpha} = P^i(\beta, \alpha, t, < TP_i, R_i, TR_i >)$, and G_α^i the goal of agent α at step i of $\mathcal{D}$. Agent α can start a negotiation dialogue at step $i+1$ of $\mathcal{D}$ if either $\neg G_\alpha^i \in TP_i$, or for all admissible arguments $(\Delta_\alpha, S_\alpha)$ of theory $T_\alpha^{\mathcal{D}_i}$ such that $(\Delta_\alpha, S_\alpha) \vdash G_\alpha^i$ there is $L \in S_\alpha$ s.t. $\neg L \in TR_i$.

For the deliberation dialogue there is no obvious definition for the initiation conditions. However we tried to capture as much as possible the intuition proposed in the literature (see e.g. [4],[8],[9]). According to this definition the shift to a deliberation dialogue happens when the participants seeking to agree upon an action or a course of action which is needed in some circumstance. In order to give a formal definition, in this paper we make the assumption, that the action to be discussed contributes to the achievement of some goal of the participants, or to the achievement of a common goal.

Definition 8. *Deliberation*
Let α and β be two agents involved in a dialogue $\mathcal{D} = L_1^{k|l} L_2^{l|k} \dots L_i^{\beta|\alpha}$. Agent α can start a deliberation dialogue on an action p, with $L_{i+1}^{\alpha|\beta} = P^{i+1}(\alpha, \beta, t, < TP_{i+1}, R_{i+1}, < TR_{i+1}^+, TR_{i+1}^- >>)$ with $p \in TP_{i+1}$, if $T_\alpha^{\mathcal{D}_i} \vdash_{GP} G$, $T_\alpha^{\mathcal{D}_i} \nvdash_{Plan} G$, $T_\alpha^{\mathcal{D}_i} \cup p \vdash G$ and $T_\alpha^{\mathcal{D}_i} \nvdash p$ and where G is a future literal.

The shift to a persuasion dialogue means that one agent disagrees with the beliefs of the other agent. The formal details are as follows. Currently in our work, persuasion is only concerned with the beliefs of agents. This is in line with the literature (see e.g. [1],[8]). However in some works persuasion is also concerned with actions. It is easy to see that a definition similar with the one proposed in the following could be proposed for the actions of agents.

Definition 9. *Persuasion*
Let α and β be two agents involved in a dialogue $\mathcal{D} = L_1^{k|l} L_2^{l|k} \dots L_i^{\beta|\alpha}$, with $L_i^{\beta|\alpha} = P^i(\beta, \alpha, t, < TP_i, < R_i^K, R_i^U >, TR_i >)$. Agent α can start a persuasion dialogue at step $i+1$ of $\mathcal{D}$, if there exists a current literal p such that $T_\alpha^{\mathcal{D}_i} \vdash_{RAC} p$ and $\neg p \in TP_i \cup R_i^K$.

A shift to an information-inquiry dialogue is similar to the shift to a deliberation dialogue, their main difference being the former concerns current literals (i.e. beliefs) while the latter actions. Informally, a shift to an information-inquiry dialogue means that one of the agents can provide to the other, part of the proof of some current literal the truth-value of which is unknown to both. This is in line with the literature (se e.g. [1],[8]).

Definition 10. *Information-Inquiry*
Let α and β be two agents involved in a dialogue $\mathcal{D} = L_1^{k|l} L_2^{l|k} \dots L_i^{\beta|\alpha}$. Agent α can start an information-inquiry dialogue at step $i+1$ of $\mathcal{D}$, on a current literal s with $L_{i+1}^{\alpha|\beta} = P^{i+1}(\alpha, \beta, t, < TP_{i+1}, < R_{i+1}^K, R_{i+1}^U >, TR_{i+1} >)$ s.t. $s \in TP_{i+1}$ and another current literal $p \in R_{i+1}^U$, if $T_\alpha^{\mathcal{D}_i} \nvdash_{RAC} s$, $T_\alpha^{\mathcal{D}_i} \cup p \vdash_{RAC} s$ and $T_\alpha^{\mathcal{D}_i} \nvdash_{RAC} p$.

This definition means that the agent α will start an information-inquiry dialogue if he searches for the truth-value of a current literal s, he knows that it can be proven by using the truth-value of the current literal p but he cannot prove p. That is why he wants to start a dialogue with another agent β who is also interested in the truth-value of s, who cannot prove s, but he can prove p.

Finally, a shift to an information-seeking dialogue is possible only if the truth-value of some current literal is unknown to one agent but known to the other. This is also in line with the literature (se e.g. [1],[8]).

Definition 11. *Information-Seeking*
Let α and β be two agents involved in a dialogue $\mathcal{D} = L_1^{k|l} L_2^{l|k} \dots L_i^{\beta|\alpha}$. Agent α can initiate an information-seeking dialogue at step $i+1$ of $\mathcal{D}$, if there exists a current literal p s.t. $T_\alpha^{\mathcal{D}_i} \nvdash_{RAC} p$, $T_\alpha^{\mathcal{D}_i} \nvdash_{RAC} \neg p$.

According to Walton [15], a dialectical shift from one dialogue type to another is *licit* if it contributes to the fulfilment of the goals of the original dialogue. If the new dialogue appears to block these goals, this shift is considered *illicit* and it is often associated with informal fallacies which are inappropriate in artificial agents dialogues. Thus, in our framework we only consider the case of licit dialectical shifts and capture this property in the following definition.

Definition 12. *Licit dialectical shift*
Let α and β be two agents participating in a dialogue $\mathcal{D}$ of type t and G_β^i the goal of agent β at step i of $\mathcal{D}$. Furthermore, let $L_i^{\alpha|\beta} = P^i(\alpha, \beta, t, < TP_i, R_i, TR_i >)$ be the locution sent by agent α to agent β at step i of $\mathcal{D}$. Agent β will initiate an embedded dialogue $\mathcal{D}'$ of type $t' \neq t$ with dialogue topic TP_{new} s.t. $TP_{new} \subseteq TP_i \cup R_i \cup TR_i$ if the following conditions hold:
1)The initiation conditions of the dialogue type t' hold
2a) $T_\beta^{\mathcal{D}_i} \cup O_{\mathcal{D}_i}^\alpha - K \nvdash g_\beta^i$ for all $g_\beta^i \subseteq G_\beta^i$, and $K = \{p | \neg p \in R_i \cup TR_i\}$ if $t' \in \{negotiation, persuasion\}$.
2b) $T_\beta^{\mathcal{D}_i} \cup O_{\mathcal{D}_i}^\alpha \nvdash g_\beta^i$ for all $g_\beta^i \subseteq G_\beta^i$ if $t' = deliberation$
3) $T_\beta^{\mathcal{D}_i} \cup TP_{new} \vdash g_\beta^i$ for some $g_\beta^i \subseteq G_\beta^i$
4) $P(\beta, \alpha, t, < TP_{new}, R_{new}, TR_{new} >)$ is not a legal locution for all possible R_{new}, TR_{new}, and $P(\beta, \alpha, t', < TP_{new}, R_{new}, TR_{new} >)$ is a legal locution for all possible R_{new}, TR_{new}.

Informally this definition says that the agent β will initiate an embedded dialogue $\mathcal{D}'$ of type t' on a new topic TP_{new} if:
1) The initiation conditions of the dialogue type t' hold

2) The agent cannot prove any goal if he removes from his knowledge the literal p whose negation belongs either to the reason or to the terms of the received locution. In the former case p can be a belief and the new dialogue will be a persuasion dialogue while in the later p can be an action or goal and the new dialogue will be a negotiation one.
3)With the new topic the agent β will be able to prove some goal
4)The locution with the new topic is not a legal locution in the current dialogue type but it is a legal locution in the new dialogue type.

Here we note that the definition of the legality of a locution depends on the adopted dialogue framework and its protocols but the exact details are beyond the scope of this paper.

Above we have not considered shifts to *information-seeking* or *information-inquiry* dialogues. For these two types of dialogues we will assume that their initiation conditions are in fact the necessary and sufficient conditions of a licit dialectical shift to them.

Finally, we define the criteria under which an agent participating in a dialogue $\mathcal{D}$ of type t accepts the request of his interlocutor to enter a new (embedded) dialogue of type t' in order to continue their discussion.

Definition 13. *Dialectical shift acceptance*
Let $\mathcal{D}$ be an open dialogue of type t between two agents α and β, and G^i_β the goal of agent β at step i of t. Furthermore, let $L_i^{\alpha|\beta} = P^i(\alpha, \beta, t', < TP_i, R_i, TR_i >)$ be the locution sent by agent α to agent β at step i of the current dialogue $\mathcal{D}$ in order to initiate an embedded dialogue $\mathcal{D}'$ of type $t' \neq t$. Agent β has to accept entering the new dialogue if the following conditions hold:
1) The initiation conditions of the dialogue type t' hold with $t' \in \{negotiation, persuasion, deliberation\}$
2) $T_\beta^{\mathcal{D}_i} \cup O^\alpha_{\mathcal{D}_i} \nvdash g^i_\beta$ for all $g^i_\beta \subseteq G^i_\beta$
3) $TP_{new} \subseteq TP_{i-1} \cup R_{i-1} \cup TR_{i-1}$ holds for the locution $L^{\beta|\alpha}_{i-1} = P^{i-1}(\beta, \alpha, t, < TP_{i-1}, R_{i-1}, TR_{i-1} >)$ sent at step $i-1$ of $\mathcal{D}$ by agent β to agent α and where $G^{i-1}_\beta \subseteq TP_{i-1}$

In the current stage of our work we consider that a dialectical shift to an *information-seeking* or *information-inquiry dialogue* is always accepted.

Our work concerns embedded dialogues among artificial agents. In this context an agreement is desirable and therefore our framework enforces agents to stay in a dialogue as long as possible by exploiting the possibility to shift among different types of dialogues according to the subject to be discussed. This is captured in condition 3 of the above definition which implies that agent β is obliged to accept a dialectical shift proposed by agent α if the proposed new topic is related to the topic, reasons or terms of the locution sent in the previous step by

himself to agent α. However, one could remove some of the above conditions or add new ones depending on the context the embedded dialogue is taking place.

The rational behind the conditions in the definitions of the licit dialectical shift and the dialectical shift acceptance, is that an agent initiates or accepts the initiation of a new type of dialogue if he expects that the outcome of the new dialogue (in case it terminates successfully) will allow the achievement of a goal which is impossible in the current dialogue. This notion of dialectical shift *efficiency* is captured formally in the following definition.

Definition 14. *Efficient dialectical shift*
Let α and β be two agents participating in a dialogue $\mathcal{D}$ of type t and G^i_α and let G^i_β be the goals of the agents at step i of $\mathcal{D}$. An embedded dialectical shift to another dialogue type $\mathcal{D}'$ will be efficient for both agents iff the following conditions hold:
1) $T^{\mathcal{D}_i}_\alpha \nvdash g^i_\alpha$ for any $g^i_\alpha \subseteq G^i_\alpha$ and $T^{\mathcal{D}_i}_\alpha \cup O^\beta_{\mathcal{D}'} \vdash g^i_\alpha$ for some $g^i_\alpha \subseteq G^i_\alpha$
2) $T^{\mathcal{D}_i}_\beta \nvdash g^i_\beta$ for any $g^i_\beta \subseteq G^i_\beta$ and $T^{\mathcal{D}_i}_\beta \cup O^\alpha_{\mathcal{D}'} \vdash g^i_\beta$ for some $g^i_\beta \subseteq G^i_\beta$

The next proposition shows that if the conditions of a licit dialectical shift hold for one of the agents, and the acceptance conditions hold for the other, the shift to the new dialogue will lead to the achievement of both agents' goals.

Proposition 1. *During an atomic dialogue $\mathcal{D}$ of type $t \in \{deliberation, negotiation\}$ between two agents, if a licit dialectical shift to another atomic dialogue $\mathcal{D}'$ of type $t' \in \{deliberation, negotiation, persuasion\}$ with $t \neq t'$ is initiated by one of the agents and accepted by the other, and the dialogue $\mathcal{D}'$ terminates successfully then it is efficient for both.*

The following proposition is a direct consequence of the way the content of a locution is defined and related to the agent's theory.

Proposition 2. *During a persuasion dialogue $\mathcal{D}$ between two agents, a licit dialectical shift to another atomic dialogue $\mathcal{D}'$ of type t' with $t' \in \{deliberation, negotiation\}$ is not possible.*

This property illustrates the fact that in the current stage of our work a persuasion dialogue can only concern the beliefs of the agents and therefore a shift to a deliberation or negotiation dialogue which may concern actions or goals is not possible.

5 Related work and conclusions

In this paper we have presented a novel approach for modelling embedded agent dialogues. Although there exists some work on the combination of atomic dialogues (see e.g. [10], [8], [14]) none of this is completely devoted to the particular study of embedded dialogues. In our work we have laid out a formal framework

based on the underlying argumentation reasoning of agents for the various issues which are necessary for the modelling of such dialogues. We have proposed a particular structure for the supporting information of the arguments exchanged between agents during a dialogue that is used to prompt (and facilitate) shifts from one dialogue type to another. We have defined the initiation conditions of the five atomic dialogues of the Walton-Krabbe typology, adopted in multi-agent context, and have shown how these are related to the supporting information. These definitions are based on a synthesis of informal descriptions proposed in the literature, but we do not pretend that these conditions may cover the totality of the possible situations. Some other works have also discussed initial conditions for the three of the five atomic dialogues (see e.g. [1], [8]) but only in a very abstract way and no direction has been given on how they could be used in the context of embedded dialogues.

Within our framework we have proposed a formal definition for the notion of licit dialectical shifts which is fundamental for the modelling of embedded dialogues along with acceptance conditions for such shifts for the participating agents. The allowed licit dialectical shifts in our framework are consistent with those of [16]. In [8] the authors have also proposed a formal framework for different atomic dialogues and have discussed issues on possible combinations. However the embedded dialogues are considered only as a case of combination of atomic dialogues with little particular attention on the formal definition of the special characteristics of such dialogues. Finally, we note that our dialogue theories for atomic and embedded dialogues can be easily implemented directly from their declarative specification in the Gorgias system [3] for argumentation and abduction.
The work presented in this paper is a first step to the formal study of embedded dialogues. Future work will concentrate on a more detailed investigation of the properties of our framework.

Acknowledgments . 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, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In *Proceedings of ICMAS98*, 1998.
2. L. Amgoud and S. Parsons. Agent dialogue with conflicting preferences. In *Proceedings of ATAL01*, 2001.
3. GORGIAS. A system for argumentation and abduction. http://www.cs.ucy.ac.cy/ nkd/gorgias, 2002.
4. D. Hitchcock, P. McBurney, and S. Parsons. A framework for deliberation dialogues. In *Proceedings of OSSA01*, 2001.
5. A. Kakas, P. Mancarella, and P. M. Dung. The acceptability semantics for logic programs. In *Proceedings of the International Conference on Logic Programming*, 1994.

6. A. Kakas, N. Maudet, and P. Moraitis. Layered strategies and protocols for argumentation-based agent interaction. In *Proceedings of ArgMAS04*, 2004.
7. A. Kakas and P. Moraitis. Argumentation based decision making for autonomous agents. In *Proceedings of AAMAS03*, 2003.
8. P. McBurney and S. Parsons. Games that agents play: a formal framework for dialogues between autonomous agents. *Journal of Logic, Language and Information*, 11:315–334, 2002.
9. P. McBurney and S. Parsons. A denotational semantics for deliberation dialogues. In *Proceedings of AAMAS04*, 2004.
10. S. Parsons, P. McBurney, and M. Wooldridge. The mechanisms of some formal inter-agent dialogue. In F.Dignum, editor, *Proceedings of Advances in Agent Communication*. Springer-Verlag, 2003.
11. S. Parsons, C. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3), 1998.
12. S. Parsons, M. Wooldridge, and L. Amgoud. An analysis of formal inter-agent dialogues. In *Proceedings of AAMAS02*, 2002.
13. S. Parsons, M. Wooldridge, and L. Amgoud. On the outcomes of formal inter-agent dialogues. In *Proceedings of AAMAS03*, 2003.
14. C. Reed. Dialogue frames in agent communication. In *Proceedings of ICMAS98*, 1998.
15. D. N. Walton. Types of dialogues, dialectical shifts and fallacies. In *Argumentation Illuminated*. 1992.
16. D.N. Walton and E.C.W. Krabbe. *Commitment in dialogue: basic concepts of interpersonal reasoning*. State University of New York Press, 1995.