An analysis of formal inter-agent dialogues

Simon Parsons and Michael Wooldridge
Department of Computer Science, University of Liverpool
Liverpool L69 7ZF, United Kingdom
{s.d.parsons,m.j.wooldridge}@csc.liv.ac.uk
Leila Amgoud
IRIT, 118 route de Narbonne, 31062 Toulouse Cedex, France
leila.amgoud@irit.fr

ABSTRACT

This paper studies argumentation-based dialogues between agents. It defines a set of locutions by which agents can trade arguments, a set of agent attitudes which relate what arguments an agent can build and what locutions it can make, and a set of protocols by which dialogues can be carried out. The paper then considers some properties of dialogues under the protocols, in particular termination and complexity, and shows how these relate to the agent attitudes.

Keywords

Agent communication, dialogue games, argumentation

1. INTRODUCTION

When building multi-agent systems, we take for granted the fact that the agents which make up the system will need to communicate. They need to communicate in order to resolve differences of opinion and conflicts of interest, work together to resolve dilemmas or find proofs, or simply to inform each other of pertinent facts. Many of these communication requirements cannot be fulfilled by the exchange of single messages. Instead, the agents concerned need to be able to exchange a sequence of messages which all bear upon the same subject. In other words they need the ability to engage in dialogues. As a result of this requirement, there has been much work on providing agents with the ability to hold such dialogues. Recently some of this work has considered argument-based approaches to dialogue, for example the work by Dignum *et al.* [5], Parsons and Jennings [15], Reed [18], Schroeder *et al.* [19] and Sycara [20].

Reed's work built on an influential model of human dialogues due to argumentation theorists Doug Walton and Erik Krabbe [21], and we also take their dialogue typology as our starting point. Walton and Krabbe set out to analyze the concept of commitment in dialogue, so as to "provide conceptual tools for the theory of argumentation" [21, page ix]. This led to a focus on persuasion dialogues, and their work presents formal models for such dialogues. In attempting this task, they recognized the need for a characterization of dialogues, and so they present a broad typology for inter-personal dialogue. They make no claims for its comprehensiveness.

Their categorization identifies six primary types of dialogues and three mixed types. The categorization is based upon: firstly, what information the participants each have at the commencement of the dialogue (with regard to the topic of discussion); secondly, what goals the individual participants have; and, thirdly, what goals are shared by the participants, goals we may view as those of the dialogue itself. As defined by Walton and Krabbe, the three types of dialogue we consider here are:

Information-Seeking Dialogues: One participant seeks the answer to some question(s) from another participant, who is believed by the first to know the answer(s).

Inquiry Dialogues: The participants collaborate to answer some question or questions whose answers are not known to any one participant.

Persuasion Dialogues: One party seeks to persuade another party to adopt a belief or point-of-view he or she does not currently hold. These dialogues begin with one party supporting a particular statement which the other party to the dialogue does not, and the first seeks to convince the second to adopt the proposition. The second party may not share this objective.

In previous work [2], we began to investigate how these different types of dialogue can be captured using a formal model of argumentation. Here we extend this work, examining some of the possible forms of information seeking, inquiry and persuasion dialogues which are possible, and identifying how the properties of these dialogues depend upon the properties of the agents engaging in them.

Note that, despite the fact that the types of dialogue we are considering are drawn from the analysis of human dialogues, we are only concerned here with dialogues between artificial agents. Unlike [9] for example, we choose to focus in this way in order to simplify our task—doing this allows us to deal with artificial languages and avoid much of the complexity inherent in natural language dialogues.

2. BACKGROUND

In this section we briefly introduce the formal system of argumentation which forms the backbone of our approach. This is inspired by the work of Dung [6] but goes further in dealing with preferences between arguments. Further details are available in [1] We start with a possibly inconsistent knowledge base Σ with no deductive closure. We assume Σ contains formulas of a propositional langage $\mathcal{L}$. $\vdash$ stands for classical inference and $\equiv$ for logical equivalence. An argument is a proposition and the set of formulae from which it can be inferred:

Definition 1. *An* argument *is a pair $A = (H, h)$ where h is a formula of $\mathcal{L}$ and H a subset of Σ such that:*

1. *H is consistent;*
2. *$H \vdash h$; and*
3. *H is minimal, so no subset of H satisfying both 1. and 2. exists.*

H is called the support *of A, written H = Support(A) and h is the* conclusion *of A written h = Conclusion(A).*

We talk of h being *supported* by the argument (H, h)

In general, since Σ is inconsistent, arguments in $\mathcal{A}(\Sigma)$, the set of all arguments which can be made from Σ, will conflict, and we make this idea precise with the notion of undercutting:

Definition 2. *Let A_1 and A_2 be two arguments of $\mathcal{A}(\Sigma)$. A_1* undercuts *A_2 iff $\exists h \in Support(A_2)$ such that $h \equiv \neg Conclusion(A_1)$.*

In other words, an argument is undercut if and only if there is another argument which has as its conclusion the negation of an element of the support for the first argument.

To capture the fact that some facts are more strongly believed[1] we assume that any set of facts has a preference order over it. We suppose that this ordering derives from the fact that the knowledge base Σ is stratified into non-overlapping sets $\Sigma_1, \ldots, \Sigma_n$ such that facts in Σ_i are all equally preferred and are more preferred than those in Σ_j where $j > i$. The preference level of a nonempty subset H of Σ, $level(H)$, is the number of the highest numbered layer which has a member in H.

Definition 3. *Let A_1 and A_2 be two arguments in $\mathcal{A}(\Sigma)$. A_1 is* preferred *to A_2 according to Pref iff $level(Support(A_1)) \leq level(Support(A_2))$.*

By $\gg^{Pref}$ we denote the strict pre-order associated with *Pref*. If A_1 is preferred to A_2, we say that A_1 is *stronger* than A_2[2]. We can now define the argumentation system we will use:

Definition 4. *An* argumentation system *(AS) is a triple $\langle \mathcal{A}(\Sigma), Undercut, Pref \rangle$ such that:*

- *$\mathcal{A}(\Sigma)$ is a set of the arguments built from Σ,*
- *Undercut is a binary relation representing defeat relationship between arguments, $Undercut \subseteq \mathcal{A}(\Sigma) \times \mathcal{A}(\Sigma)$, and*
- *Pref is a (partial or complete) preordering on $\mathcal{A}(\Sigma) \times \mathcal{A}(\Sigma)$.*

The preference order makes it possible to distinguish different types of relation between arguments:

Definition 5. *Let A_1, A_2 be two arguments of $\mathcal{A}(\Sigma)$.*

- *If A_2 undercuts A_1 then A_1* defends itself *against A_2 iff $A_1 \gg^{Pref} A_2$. Otherwise, A_1* does not defend itself*.*
- *A set of arguments $\mathcal{S}$* defends *A iff: $\forall$ B undercuts A and A does not defend itself against B then $\exists$ $C \in \mathcal{S}$ such that C undercuts B and B does not defend itself against C.*

Henceforth, $C_{Undercut,Pref}$ will gather all non-undercut arguments and arguments defending themselves against all their undercutting arguments. In [1], it was shown that the set $\underline{\mathcal{S}}$ of acceptable arguments of the argumentation system $\langle \mathcal{A}(\Sigma), Undercut, Pref \rangle$ is the least fixpoint of a function $\mathcal{F}$:

$$\begin{aligned} \mathcal{S} &\subseteq \mathcal{A}(\Sigma) \\ \mathcal{F}(\mathcal{S}) &= \{(H,h) \in \mathcal{A}(\Sigma) | (H,h) \text{ is defended by } \mathcal{S}\} \end{aligned}$$

Definition 6. *The set of* acceptable *arguments for an argumentation system $\langle \mathcal{A}(\Sigma), Undercut, Pref \rangle$ is:*

$$\begin{aligned} \underline{\mathcal{S}} &= \bigcup \mathcal{F}_{i \geq 0}(\emptyset) \\ &= C_{Undercut,Pref} \cup \left[\bigcup \mathcal{F}_{i \geq 1}(C_{Undercut,Pref}) \right] \end{aligned}$$

An argument is acceptable *if it is a member of the acceptable set.*

An acceptable argument is one which is, in some sense, proven since all the arguments which might undermine it are themselves undermined.

3. LOCUTIONS

As in our previous work [2, 4], agents use the argumentation mechanism described above as a basis for their reasoning and their dialogues. Agents decide what they themselves know by determining which propositions they have acceptable arguments for. They trade propositions for which they have acceptable arguments, and accept propositions put forward by other agents if they find that the arguments are acceptable. The exact locutions and the way that they are exchanged define a formal dialogue game which agents engage in.

Dialogues are assumed to take place between two agents, P and C[3]. Each agent has a knowledge base, Σ_P and Σ_C respectively, containing their beliefs. In addition, each agent has a further knowledge base, accessible to both agents, containing commitments made in the dialogue. These commitment stores are denoted $CS(P)$ and $CS(C)$ respectively, and in this dialogue system (unlike that of [4] for example) an agent's commitment store is just a subset of its knowledge base. Note that the union of the commitment

[1] Here we only deal with beliefs, though the approach can also handle desires and intentions as in [16] and could be extended to cope with other mental attitudes.

[2] We acknowledge that this model of preferences is rather restrictive and in the future intend to work to relax it.

[3] The names stemming from the study of persuasion dialogues—P argues "pro" some proposition, and C argues "con".

stores can be viewed as the state of the dialogue at a given time. Each agent has access to their own private knowledge base and both commitment stores. Thus P can make use of $\langle \mathcal{A}(\Sigma_P \cup CS(C)), \mathit{Undercut}, \mathit{Pref}\rangle$[4] and C can make use of $\langle \mathcal{A}(\Sigma_C \cup CS(P), \mathit{Undercut}, \mathit{Pref}\rangle$.

All the knowledge bases contain propositional formulas and are not closed under deduction, and all are stratified according to degree of belief as discussed above. Here we assume that these degrees of belief are static and that both the players agree on them, though it is possible [3] to combine different sets of preferences, and it is also possible to have agents modify their beliefs on the basis of the reliability of their acquaintances [14].

With this background, we can present the set of dialogue moves that we will use. For each move, we give what we call rationality rules and update rules. These are based on the rules suggested by [11]. The rationality rules specify the preconditions for playing the move. Unlike those in [2, 4] these are not absolute, but are defined in terms of the agent attitudes discussed in Section 4. The update rules specify how commitment stores are modified by the move.

In the following, player P adresses the move to player C. We start with the assertion of facts:

assert(p). where p is a propositional formula.

> **rationality** the usual assertion condition for the agent.
>
> **update** $CS_i(P) = CS_{i-1}(P) \cup \{p\}$ and $CS_i(C) = CS_{i-1}(C)$

Here p can be any propositional formula, as well as the special character $\mathcal{U}$, discussed below.

assert(S). where S is a set of formulas representing the support of an argument.

> **rationality** the usual assertion condition for the agent.
>
> **update** $CS_i(P) = CS_{i-1} \cup S$ and $CS_i(C) = CS_{i-1}(C)$

The counterpart of these moves are the acceptance moves:

accept(p). p is a propositional formula.

> **rationality** The usual acceptance condition for the agent.
>
> **update** $CS_i(P) = CS_{i-1}(P) \cup \{p\}$ and $CS_i(C) = CS_{i-1}(C)$

accept(S). S is a set of propositional formulas.

> **rationality** the usual acceptance condition for every $s \in S$.
>
> **update** $CS_i(P) = CS_{i-1}(P) \cup S$ and $CS_i(C) = CS_{i-1}(C)$

There are also moves which allow questions to be posed.

[4]Which, of course, is the same as $\langle \mathcal{A}(\Sigma_P \cup CS(P) \cup CS(C)), \mathit{Undercut}, \mathit{Pref}\rangle$.

challenge(p). where p is a propositional formula.

> **rationality** $\emptyset$
>
> **update** $CS_i(P) = CS_{i-1}(P)$ and $CS_i(C) = CS_{i-1}(C)$

A challenge is a means of making the other player explicitly state the argument supporting a proposition. In contrast, a question can be used to query the other player about any proposition.

question(p). where p is a propositional formula.

> **rationality** $\emptyset$
>
> **update** $CS_i(P) = CS_{i-1}(P)$ and $CS_i(C) = CS_{i-1}(C)$

We refer to this set of moves as the set $\mathcal{M}'_{\mathcal{DC}}$ since they are a variation on the set $\mathcal{M}_{\mathcal{DC}}$ from [2]—the main difference from the latter is that there are no "dialogue conditions". Instead we explicitly define the protocol for each type of dialogue in Section 5. The locutions in $\mathcal{M}'_{\mathcal{DC}}$ are similar to those discussed in legal reasoning [7, 17] and it should be noted that there is no *retract* locution. Note that these locutions are ones used within dialogues. Further locutions such as those discussed in [13] would be required to frame dialogues.

4. AGENT ATTITUDES

One of the main aims of this paper is to explore how the kinds of dialogue in which agents engage depends upon features of the agents themselves (as opposed, for instance, to the kind of dialogue in which the agents are engaged or the information in the knowledge-bases of the agents). In particular, we are interested in the effect of these features on the way in which agents determine what locutions can be made within the confines of a given dialogue protocol through the application of differing rationality conditions.

As is clear from the definition of the locutions, there are two different kinds of rationality conditions—one which determines if something may be asserted, and another which determines whether something can be accepted. The former we call assertion conditions, the latter we call acceptance conditions and talk of agents having different attitudes which relate to particular conditions.

Definition 7. *An agent may have one of two* assertion *attitudes.*

- *a* confident *agent can assert any proposition p for which it can construct an argument (S, p).*
- *a* thoughtful *agent can assert any proposition p for which it can construct an acceptable argument (S, p).*

Thus a thoughtful agent will only put forward propositions which, so far as it knows, are correct. A confident agent won't stop to check that this is the case. It might seem worthwhile also defining what we might call a *thoughtless* agent, which can assert any proposition which is either in, or may be inferred from, its knowledge base, but it is easy to show that:

Proposition 1. *The set of non-trivial propositions which can be asserted by a thoughtless agent using an argumentation system $\langle \mathcal{A}(\Sigma), \mathit{Undercut}, \mathit{Pref}\rangle$ is exactly the set which can be asserted by a confident agent using the same argumentation system.*

PROOF. *Consider a confident agent G and a thoughtless agent H with the same argumentation system. G can assert exactly those propositions that it has an argument for. So by Definition 1 it can assert any p which it can infer from a minimal consistent subset of Σ, including all the propositions q in Σ (these are the conclusions of the arguments $(\{q\}, q)$). H can assert any proposition which is either in Σ (which will be exactly the same as those G can assert) or can be infered from it. Those propositions which are non-trivial will be those that can be inferred from a consistent subset of Σ. These latter will clearly be ones for which an argument can be built, and so exactly those that can be asserted by G.* □

Thus the idea of a thoughtless agent adds nothing to our classification.

At the risk of further overloading some well-used terms we can define acceptance conditions:

DEFINITION 8. *An agent may have one of three* acceptance *attitudes.*

- *a* credulous *agent can accept any proposition p if it is backed by an argument.*
- *a* cautious *agent can accept any proposition p if it is unable to construct a stronger argument for $\neg p$.*
- *a* skeptical *agent can accept any proposition p if there is an acceptable argument for p.*

With a pair of agents that are thoughtful and skeptical, we recover the rationality conditions of the dialogue system in [2].

Clearly skeptical agents are more demanding than credulous ones in terms of the conditions they put on accepting information. Typically, a skeptical agent which is presented with an assertion of p will challenge p to obtain the argument for it, and then validate that this argument is acceptable given what it knows. We can consider even more demanding agents. For example, we can imagine a *querelous* agent which will only accept a proposition if it can not only validate the acceptability of the argument for that proposition, but also the acceptability of arguments for all the propositions in that argument, and all the propositions in those arguments, and so on.

However, it turns out that:

PROPOSITION 2. *The set of propositions acceptable to a skeptical agent using an argumentation system $\langle \mathcal{A}(\Sigma)$, Undercut, Pref$\rangle$ is exactly the same as the set of propositions acceptable to a querelous agent using the same argumentation system.*

PROOF. *Consider a thoughtful agent G and a querelous agent H with the same argumentation system. By definition, G can accept any proposition p whose support S is either not attacked by any argument which is built from Σ, or is defended by an argument which is part of the acceptable set of $\mathcal{A}(\Sigma)$. In other words, G will only accept p if all the $s \in S$ are themselves supported by acceptable arguments (which might just be $(\{s\}, s)$ if there is no argument for $\neg s$). This is exactly the set of conditions under which H will accept p.* □

In other words once we require an argument to be acceptable, we also require that any proposition which is part of the support for that argument is also acceptable. Thus the notion of a querelous agent adds nothing to our classification.

5. DIALOGUE TYPES

With the agent attitudes specified, we can begin to look at different types of dialogue in detail giving protocols for each. These protocols are intentionally simple, to make it possible to provide a detailed analysis of them as a baseline from which more complex protocols can be examined. An important feature common to all these protocols is that no agent is allowed to repeat a locution. If this prevents the agent from making any locution, the dialogue terminates.

5.1 Information-seeking

In an information seeking dialogue, one participant seeks the answer to some question from another participant. If the information seeker is agent A and the other agent is B, then we can define the protocol $\mathcal{IS}$ for an information seeking dialogue about a proposition p as follows:

1. A asks *question*(p).
2. B replies with either *assert*(p), *assert*$(\neg p)$, or *assert*$(\mathcal{U})$. Which will depend upon the contents of its knowledge-base and its assertion attitude. $\mathcal{U}$ indicates that, for whatever reason B cannot give an answer.
3. A either *accept*s B's response, if its acceptance attitude allows, or *challenge*s. $\mathcal{U}$ cannot be *challenge*d and as soon as it is asserted, the dialogue terminates without the question being resolved.
4. B replies to a *challenge* with an *assert*(S), where S is the support of an argument for the last proposition challenged by A.
5. Go to 3 for each proposition in S in turn.

Note that A accepts whenever possible, only being able to challenge when unable to accept—"only" in the sense of only being able to challenge then and *challenge* being the only locution other than *accept* that it is allowed to make. More flexible dialogue protocols are allowed, as in [2], but at the cost of possibly running forever[5].

There are a number of interesting properties that we can prove about this protocol, some of which hold whatever acceptance and assertion attitudes the agents have, and some of which are more specific. We have:

PROPOSITION 3. *When subject to challenge(p) for any p it has asserted, a confident or thoughtful agent G can always respond.*

PROOF. *In order to respond to a challenge(p), the agent has to be able to produce an argument (S, p). Since, by definition, both confident and thoughtful agents only assert propositions for which they have arguments, these arguments can clearly be produced if required. This holds even for the propositions in S. For a proposition to be in S by Definition 1 it must be part of a consistent, minimal subset of Σ_G which entails p. Any such proposition q is the conclusion of an argument $(\{q\}, q)$ and this argument is easily generated.* □

This first result ensures that step 4 can always follow from step 3, and the dialogue will not get stuck at that point.

[5]The protocol in [2] allows an agent to interject with *question*(p) for any p at virtually any point, allowing the dialogue to be fillibusted by issuing endless *question*s about arbitrary formulae.

It also leads to another result—since with this protocol our agents only put forward propositions which are backed by arguments, a credulous agent would have to accept any propostion asserted by an agent:

PROPOSITION 4. *A credulous agent G operating under protocol $\mathcal{IS}$ will always accept a proposition asserted by a confident or thoughtful agent H.*

PROOF. *When H asserts p, G will initially challenge it (for p to be acceptable it must be backed by an argument, but no argument has been presented by H and if G had an argument for p it would not have engaged in the information seeking dialogue). By Proposition 3, H will always be able to generate such an argument, and by the definition of its acceptance condition and the protocol $\mathcal{IS}$, G will then accept it.* □

This result is crucial in showing that if A is a credulous agent, then the dialogue will always terminate, but what if it is more demanding? Well, it turns out that:

PROPOSITION 5. *An information-seeking dialogue under protocol $\mathcal{IS}$ between a credulous, cautious or skeptical agent G and a confident or thoughtful agent H will always terminate.*

PROOF. *At step 2. of the protocol H either replies with p, $\neg p$ or $\mathcal{U}$. If it is $\mathcal{U}$, the dialogue terminates. G then considers p. If G is creduous, then by Corollary 4, G will accept the proposition and the dialogue will terminate.*

If G is cautious, then at step 3, it will either accept p, or have an argument for $\neg p$. In the former case the dialogue terminates immediately. In the latter case G will challenge p and by Proposition 3 receive the support S. If G doesn't have an argument against any of the $s \in S$, then they will be accepted, but this will not make G accept p. The only locution that G could utter is challenge(p), but it is prevented from doing this, and the dialogue terminates. If G does have an argument for the negation of any of the $s \in S$, then it will challenge them. As in the proof of Proposition 3 this will produce an argument $(\{s\}, s)$ from H, and G will not be able to accept this. It also cannot challenge this since this would repeat its challenge of s, and the dialogue will terminate.

If G is skeptical, then the process will be very similar. At step 3, G will not be able to accept p (for the same kind of reason as in the proof of Proposition 4), so will challenge it and receive the support S. This support may mean that G has an acceptable argument for p in which case the dialogue terminates. If this argument is not acceptable, then G will challenge the $s \in S$ for which it has an undercutting argument. Again, this will produce an argument $(\{s\}, s)$ from H which won't make the argument for p acceptable. G cannot make any further locutions, and the dialogue will terminate. □

While this result is a good one, because of the guarantee of termination, the proof illustrates a limitation of the dialogue protocol.

Whether G is skeptical or cautious, it will either immediately accept p or never accept it whatever H says. That is H will never persuade G to change its mind. The reason for this is that the dialogue protocol neither makes G assert into $CS(G)$ the grounds for not accepting p (thus giving H the opportunity to attack the relevant argument), nor gives H the chance to do anything other than assert arguments which support p.

This position can be justified since $\mathcal{IS}$ is intended only to capture information seeking. If we want H to be able to persuade G, then the agents are engaging in a persuasion dialogue, albeit one that is embedded in an information seeking dialogue as in [13], and this case is thus dealt with below.

5.2 Inquiry

In an inquiry dialogue, the participants collaborate to answer some question whose answer is not known to either. There are a number of ways in which one might construct an inquiry dialogue (for example see [12]). Here we present one simple possibility. We assume that two agents A and B have already agreed to engage in an inquiry about some proposition p by some control dialogue as suggested in [13], and from this point can adopt the following protocol $\mathcal{I}$:

1. A asserts $q \rightarrow p$ for some q or $\mathcal{U}$.
2. B accepts $q \rightarrow p$ if its acceptance attitude allows, or challenges it.
3. A replies to a *challenge* with an *assert*(S), where S is the support of an argument for the last proposition challenged by B.
4. Goto 2 for each proposition $s \in S$ in turn, replacing $q \rightarrow p$ by s.
5. B *asserts* q, or $r \rightarrow q$ for some r, or $\mathcal{U}$.
6. If $A(CS(A) \cup CS(B))$ includes an argument for p which is acceptable to both agents, then the dialogue terminates successfully.
7. Go to 5, reversing the roles of A and B and substituting r for q and some t for r.

This protocol is basically a series of implied $\mathcal{IS}$ dialogues. First A asks "do you know of anything which would imply p were it known?". B replies with one, or the dialogue terminates with $\mathcal{U}$. If A accepts the implication, B asks "now, do you know q, or any r which would imply q were it known?", and the process repeats until either the process bottoms out in a proposition which both agents agree on, or there is no new implication to add to the chain. Because of this structure, it is easy to show that:

PROPOSITION 6. *An inquiry dialogue $\mathcal{I}$ between two agents G and H with any acceptance and assertion attitudes will terminate.*

PROOF. *The dialogue starts with an implied $\mathcal{IS}$ dialogue. By Proposition 5 this dialogue will terminate. If it terminates with a result other than $\mathcal{U}$, then it is followed with a second $\mathcal{IS}$ dialogue in which the roles of the agents are reversed. Again by Proposition 5 this dialogue will terminate, possibly with a proof that is acceptable to both agents. If this second dialogue does not end with a proof or a $\mathcal{U}$, then it is followed with another $\mathcal{IS}$ dialogue in which the roles of the agents are again reversed. This third dialogue runs just like the second. The iteration will continue until either one of the agents responds with a $\mathcal{U}$, or the chain of implications is ended. One or other will happen since the agents can only build a finite number of arguments (since* *arguments have supports which are minimal consistent sets of the finite knowledge base), and agents are not allowed to repeat themselves. When the iteration terminates, so does the dialogue.* □

However, it is also true that this rather rigid protocol may prevent a proof being found even though one is available to the agents if they were to make a different set of assertions. More precisely, we have:

PROPOSITION 7. *Two agents G and H which engage in a inquiry dialogue for p, using protocol $\mathcal{I}$, may find the dialogue terminates unsuccessfully even when $A(\Sigma_G \cup \Sigma_H)$ provides an argument p which both agents would be able to accept.*

PROOF. *Consider G has $\Sigma_G = \{q \rightarrow p, r \rightarrow p\}$ and H has $\Sigma_H = \{r\}$. Clearly together both agents can produce $(\{r, r \rightarrow p\}, p)$, and this will be acceptable to both agents no matter their acceptance attitude, but if G starts by asserting $q \rightarrow p$ the agents will never find this proof.* □

Of course, it is possible to design protocols which don't suffer from this problem, by allowing an agent to assert all the $r \rightarrow q$ which are relevant at any point in the dialogue (turning the dialogue into a breadth-first search for a proof rather than a depth first one) or by allowing the dialogue to backtrack.

Another thing to note is that, in contrast to the information seeking dialogue, in inquiry dialogues the relationship between the agents is symmetrical in the sense that both are asserting and accepting arguments. Thus both an agent's assertion attitude and acceptance attitude come into play. As a result, in the case of a confident but skeptical agent, it is possible for an agent to assert an argument that it would not find acceptable itself. This might seem odd at first, but on reflection seems more reasonable (consider the kind of inquiry dialogue one might have with a child), not least when one considers that a confident assertion attitude can be seen as one which responds to resource limitations—assert something that seems reasonable and only look to back it up if there is a reason (its unacceptability to another agent) which suggests that it is problematic.

5.3 Persuasion

In a persuasion dialogue, one party seeks to persuade another party to adopt a belief or point-of-view he or she does not currently hold. The dialogue game DC, on which the moves in [2] are based, is fundamentally a persuasion game, so the protocol below results in games which are very like those described in [2]. This protocol, $\mathcal{P}$, is as follows, where agent A is trying to persuade agent B to accept p.

1. A *assert*s p.
2. B *accept*s p if its acceptance attitude allows, if not B *assert*s $\neg p$ if it is allowed to, or otherwise *challenge*s p.
3. If B asserts $\neg p$, then goto 2 with the roles of the agents reversed and $\neg p$ in place of p.
4. If B has *challenge*d, then:
   (a) A asserts S, the support for p;
   (b) Goto 2 for each $s \in S$ in turn.

If at any point an agent cannot make the indicated move, it has to concede the dialogue game. If A concedes, it fails to persuade B that p is true. If B concedes, then A has succeeded in persuading it. An agent also concedes the game if at any point if there are no propositions made by the other agent that it hasn't accepted.

Once again the form of this dialogue has much in common with inquiry dialogues. The dialogue starts as if B has asked A if p is true, and A's response is handled in the same way as in an inquiry unless B has a counter-argument in which case it can assert it. This assertion is like spinning off a separate $\mathcal{IS}$ dialogue in which A asks B if $\neg p$ is true. Since we already have a termination result for $\mathcal{IS}$ dialogues, it is simple to show that:

PROPOSITION 8. *A persuasion dialogue under protocol $\mathcal{P}$ between two agents G and H will always terminate.*

PROOF. *A dialogue under $\mathcal{P}$ is just like an information seeking dialogue under $\mathcal{IS}$ in which agents are allowed to reply to the assertion of a proposition p with the assertion of $\neg p$ as well as the usual responses. Since we know that a dialogue under $\mathcal{IS}$ always terminates, it suffices to show that the assertion of $\neg p$ does not lead to non-termination. Since the only difference between the sub-dialogue spawned by the assertion of $\neg p$ and a $\mathcal{IS}$ dialogue is the possibility of the agent to which $\neg p$ is asserted asserting p in response, then this is the only way in which non-termination can occur. However, this assertion of p is not allowed since it would repeat the assertion that provoked the $\neg p$ and so the dialogue would terminate. Thus a $\mathcal{P}$ dialogue will always terminate.* □

Again there is some symmetry between the agents, but there is also a considerable asymmetry which stems from the fact that A is effectively under a burden of proof so it has to win the argument in order to convince B, while B just has to fail to lose to not be convinced. Thus if A and B are both confident-cautious and one has an argument for p and the other has one for $\neg p$, and neither argument is stronger than the other, despite the fact that the arguments "draw", A will lose the exchange and B will not be convinced. This is exactly the same kind of behaviour that is exhibited by all persuasion dialogues in the literature.

6. COMPLEXITY OF DIALOGUES

Having examined some of the properties of the dialogues, we consider their computational complexity. Since the protocols are based on reasoning in logic we know that the complexity will be high—our aim in this analysis is to establish exactly where the complexity arises so that we can try and reduce it,for example, as in [22], but suitable choice of language.

To study this issue, we return to Definition 1. Given a knowledge base Σ, we will say there is a *prima facie* argument for a particular conclusion h if $\Sigma \vdash h$, i.e., if it is possible to prove the conclusion from the knowledge base. The existence of a prima facie argument does not imply the existence of a "usable" argument, however, as Σ may be inconsistent. Since establishing proof in propositional logic is co-NP-complete, we can immediately conclude:

PROPOSITION 9. *Given a knowledge base Σ and a conclusion h, determining whether there is a prima facie argument for h from Σ is co-NP-complete.*

We will say a pair (H, h) is a consistent prima facie argument over Σ if H is a consistent subset of Σ and $H \vdash h$. Determining whether or not there is a consistent prima facie argument for some conclusion is immediately seen to be harder.

PROPOSITION 10. *Given a knowledge base Σ and conclusion h, determining whether there is a consistent prima facie argument for h over Σ is Σ_2^p-complete.*

PROOF. *The following Σ_2^p algorithm decides the problem:*

1. *Existentially guess a subset H of Σ together with a valuation v for H.*
2. *Verify that $v \models H$.*
3. *Universally select each valuation v' of H, and verify that $v' \models H \rightarrow h$.*

The algorithm has two alternations, the first being an existential, the second a universal, and so it is indeed a Σ_2^p algorithm. The existential alternation involves guessing a support for h together with a witness to the consistency of this support. The universal alternation verifies that $H \rightarrow h$ is valid, and so $H \vdash h$. Thus the problem is in Σ_2^p.

To show the problem is Σ_2^p-hard, we do a reduction from the QBF$_{2,\exists}$ *problem [10, p96]. An instance of* QBF$_{2,\exists}$ *is given by a quantified boolean formula with the following structure:*

$$\exists x_1, \ldots, x_k \, \forall y_1, \ldots, y_l \;\; \chi \tag{1}$$

where χ is a propositional logic formula over Boolean variables $x_1, \ldots, x_k, y_1, \ldots, y_l$. Such a formula is true if there are values we can give to $x_1, \ldots, x_k$, such that for all values we can give to $y_1, \ldots, y_l$, the formula χ is true. Here is an example of such a formula.

$$\exists x_1 \forall x_2 [(x_1 \vee x_2) \wedge (x_1 \vee \neg x_2)] \tag{2}$$

Formula (2) in fact evaluates to true. (If x_1 is true, then for all values of x_2, the overall formula is true.)

Given an instance (1) of QBF$_{2,\exists}$*, we define the conclusion h to be $h = \chi$, and then define the knowledge base Σ as*

$$\Sigma = \{x_1 \leftrightarrow \bot, x_1 \leftrightarrow \top, \ldots, x_k \leftrightarrow \bot, x_k \leftrightarrow \top\}.$$

where $\top$ and $\bot$ are logical constants for truth and falsehood respectively. Any consistent subset of Σ defines a consistent partial valuation for the body of (1); variables not given a valuation by a subset are assumed to be "don't care". We claim that input formula (1) is true iff there exists a consistent prima facie argument for h given knowledge base Σ. Intuitively, in considering subsets of Σ, we are actually examining all values that may be assigned to the existentially quantified variables $x_1, \ldots, x_k$. Since the reduction is clearly polynomial time, we are done. □

Now, knowing that there exists a consistent prima facie argument for conclusion h over Σ implies the existence of a *minimal* argument for h over Σ (although it does not tell us what this minimal argument is). We can thus conclude:

COROLLARY 1. *Given a knowledge base Σ and conclusion h, determining whether there is an argument for h (i.e., a minimal consistent prima facie argument for h — Definition 1) over Σ is Σ_2^p-complete.*

The next obvious question is as follows: given (H, h), where $H \vdash h$, is it minimal?

COROLLARY 2. *Given a knowledge base Σ and prima facie argument (H, h) over Σ, the problem of determining whether (H, h) is minimal is Π_2^p-complete.*

PROOF. *For membership of Π_2^p, consider the following Σ_2^p algorithm, which decides the complement of the problem:*

1. *Existentially select a subset H' of H and a valuation v for H'.*
2. *Verify that $v \models H'$.*
3. *Universally select each valuation v' for H'.*
4. *Verify that $v' \models H' \rightarrow h$.*

The algorithm contains two alternations, an existential followed by an universal, and so is indeed a Σ_2^p algorithm. The algorithm works by guessing a subset H' of H, showing that this subset is consistent, and then showing that $H' \rightarrow h$ is a tautology, so $H' \vdash h$. Since the complement of the problem under consideration is in Σ_2^p, and co-$\Sigma_2^p = \Pi_2^p$, it follows that the problem is in Π_2^p.

To show completeness, we reduce the QBF$_{2,\exists}$ *to the complement of the problem, i.e., to showing that an argument is not minimal. If an argument (H, h) is not minimal, then there will exist some consistent subset H' of H such that $H' \vdash h$. The reduction is identical to that above: we set $H = \{x_1 \leftrightarrow \bot, x_1 \leftrightarrow \top, \ldots, x_k \leftrightarrow \bot, x_k \leftrightarrow \top\}$ and set $h = \chi$. We then ask whether there is a consistent subset H' of H such that $H' \vdash h$. Since we have reduced a Σ_2^p-complete problem to the complement of the problem under consideration, it follows that the problem is Π_2^p-hard.* □

These results allow us to handle the complexity of dialogues involving confident, credulous and cautious agents, which are only interested in whether arguments can be built for given propositions. For thoughtful and skeptical agents we need to consider whether an argument is undercut.

PROPOSITION 11. *Given a knowledge base Σ and an argument (H, h) over Σ, the problem of showing that (H, h) has an undercutter is Σ_2^p-complete.*

PROOF. *The following Σ_2^p algorithm decides this problem:*

1. *Existentially guess (i) a subset H' of Σ; (ii) a support formula $h' \in H$ to undercut; and (iii) a valuation v.*
2. *Verify that $v \models H'$.*
3. *Universally select each valuation v' of H'.*
4. *Verify that (i) $v' \models H' \rightarrow h'$ and (ii) $v' \models \neg h \leftrightarrow h'$.*

For hardness, there is a straightforward reduction from the QBF$_{2,\exists}$ *problem, essentially identical to the reductions given in proofs above — we therefore omit it.* □

As a corollary, the problem of showing that (H, h) has *no* undercutter is Π_2^p-complete.

These results are sufficient to demonstrate the worst-case intractability of argumentation-based approaches for skepical and thoughtful agents using propositional logic. They thus motivate the investigation of the behaviour of agents with different attitudes and the use of other logics. These matters are explored in an extended version of this paper.

7. CONCLUSIONS

This paper has examined three types of argumentation-based dialogue between agents—information seeking, inquiry and persuasion, from the typology of [21]—defining a precise protocol for each and examining some important properties of that protocol. In particular we have shown that each protocol leads to dialogues that are guaranteed to terminate, and we have considered some aspects of the complexity of these dialogues. The exact form of the dialogues depends on what messages agents send and how they respond to messages they receive. This aspect of the dialogue is not specified by the protocol, but by some decision-making apparatus in the agent. Here we have considered this decision to be determined by the agents' attitude, and we have shown how this attitude affects their behaviour in the dialogues they engage in.

Both of these aspects extend previous work in this field. In particular, they extend the work of [2] by precisely defining a set of protocols (albeit quite rigid ones) and a range of agent attitudes—in [2] only one protocol, for persuasion, and only one attitude, broadly thoughtful-skeptical, were considered.

More work, of course, remains to be done in this area. Particularly important is determining the relationship between the locutions we use in these dialogues and those of agent communication languages such as the FIPA ACL, examining the effect of adding new locutions (such as *retract*) to the language, and identifying additional properties of the dialogues (such as whether the order in which arguments are made affects the outcome of the dialogue). We are currently investigating these matters along with further dialogue types, more complex kinds of the dialogue types studied here, such planning dialogues [8], and additional complexity issues (including the effect of languages other than propositional logic).

Acknowledgments

This work was partly funded by the EU funded Project IST-1999-10948.

8. REFERENCES

[1] L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation framework. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 1–7, 1998.

[2] L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In E. Durfee, editor, *Proceedings of the Fourth International Conference on Multi-Agent Systems*, pages 31–38, Boston, MA, USA, 2000. IEEE Press.

[3] L. Amgoud and S. Parsons. Agent dialogues with conflicting preferences. In J.-J. Meyer and M. Tambe, editors, *Proceedings of the 8th International Workshop on Agent Theories, Architectures and Languages*, pages 1–15, 2001.

[4] L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proceedings of the Fourteenth European Conference on Artificial Intelligence*, pages 338–342, Berlin, Germany, 2000. IOS Press.

[5] F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Agent theory for team formation by dialogue. In C. Castelfranchi and Y. Lespérance, editors, *Seventh Workshop on Agent Theories, Architectures, and Languages*, pages 141–156, Boston, USA, 2000.

[6] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.

[7] T. F. Gordon. The pleadings game. *Artificial Intelligence and Law*, 2:239–292, 1993.

[8] B. J. Grosz and S. Kraus. The evolution of sharedplans. In M. J. Wooldridge and A. Rao, editors, *Foundations of Rational Agency*, volume 14 of *Applied Logic*. Kluwer, The Netherlands, 1999.

[9] B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

[10] D. S. Johnson. A catalog of complexity classes. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science Volume A: Algorithms and Complexity*, pages 67–161. Elsevier, 1990.

[11] N. Maudet and F. Evrard. A generic framework for dialogue game implementation. In *Proceedings of the 2nd Workshop on Formal Semantics and Pragmatics of Dialogue*, University of Twente, The Netherlands, May 1998.

[12] P. McBurney and S. Parsons. Risk agoras: Dialectical argumentation for scientific reasoning. In C. Boutilier and M. Goldszmidt, editors, *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, Stanford, CA, USA, 2000. UAI.

[13] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language, and Information*, (to appear).

[14] S. Parsons and P. Giorgini. An approach to using degrees of belief in BDI agents. In B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh, editors, *Information, Uncertainty, Fusion*. Kluwer, Dordrecht, 1999.

[15] S. Parsons and N. R. Jennings. Negotiation through argumentation — a preliminary report. In *Proceedings of Second International Conference on Multi-Agent Systems*, pages 267–274, 1996.

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

[17] H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127:187–219, 2001.

[18] C. Reed. Dialogue frames in agent communications. In Y. Demazeau, editor, *Proceedings of the Third International Conference on Multi-Agent Systems*, pages 246–253. IEEE Press, 1998.

[19] M. Schroeder, D. A. Plewe, and A. Raab. Ultima ratio: should Hamlet kill Claudius. In *Proceedings of the 2nd International Conference on Autonomous Agents*, pages 467–468, 1998.

[20] K. Sycara. Argumentation: Planning other agents' plans. In *Proceedings of the Eleventh Joint Conference on Artificial Intelligence*, pages 517–523, 1989.

[21] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning.*

State University of New York Press, Albany, NY, 1995.
[22] M. Wooldridge and S. Parsons. Languages for negotiation. In W. Horn, editor, *Proceedings of the Fourteenth European Conference on Artificial Intelligence*, pages 393– 397, Berlin, Germany, 2000. IOS Press.