Properties and complexity of some formal inter-agent dialogues*

Simon Parsons
Department of Computer and Information Science
Brooklyn College, City University of New York
2900 Bedford Avenue, Brooklyn, NY 11210, USA.
parsons@sci.brooklyn.cuny.edu

Michael Wooldridge
Department of Computer Science, University of Liverpool
Liverpool L69 7ZF, United Kingdom
m.j.wooldridge@csc.liv.ac.uk

Leila Amgoud
IRIT, 118 route de Narbonne, 31062 Toulouse Cedex, France
leila.amgoud@irit.fr

February 3, 2003

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, dialogue outcomes, and complexity, and shows how these relate to the agent attitudes.

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

*This is a revised and expanded version of [31].

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 — for example by Dignum and colleagues [8, 9], Grosz and Kraus [15], Parsons and Jennings [28, 30], Reed [37], Schroeder *et al.* [40], and Sycara [42].

Reed's work built on an influential model of human dialogues due to argumentation theorists Doug Walton and Erik Krabbe [45], 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" [45, 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 six primary dialogue types are (re-ordered from [45]):

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.

Negotiation Dialogues: The participants bargain over the division of some scarce resource in a way acceptable to all, with each individual party aiming to maximize his or her share. The goal of the dialogue may be in conflict with the individual goals of each of the participants.[1]

Deliberation Dialogues: Participants collaborate to decide what course of action to take in some situation. Participants share a responsibility to decide the course of action, and either share a common set of intentions or a willingness to discuss rationally whether they have shared intentions.

[1] Note that this definition of negotiation is that of Walton and Krabbe. Arguably negotiation dialogues may involve other issues besides the division of scarce resources.

Eristic Dialogues: Participants quarrel verbally as a substitute for physical fighting, with each aiming to win the exchange. We include Eristic dialogues here for completeness, but we do not discuss them further.

In previous work [3, 6], 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 that 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 [16] 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. This issue is discussed in more depth in Section 8.

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 [11] 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 language $\mathcal{L}$. $\vdash$ 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[2] we assume that any set of facts has a preference order over it (other approaches to quantifying belief, such as probability, could also be used in conjunction with our approach). 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.

This is clearly a very restricted notion of how to handle preferences. Other approaches to handling preferences, which could also be used along with our approach, are surveyed in [10], and we discuss how this argumentation system can be extended with a more flexible notion of preferences in [5] (which also allows arguments for and against preferences themselves). We stick with the simple model here for ease of exposition, noting that nothing in the rest of the paper hinges upon it—what is required for the argumentation system is a notion of preference, the use made of this notion does not depend upon the way that preferences are represented.

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 the 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.

[2] Here we only deal with beliefs, though the approach can also handle desires and intentions [6, 30] and could be extended to cope with other mental attitudes.

- *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 [2], it was shown that the set $\underline{\mathcal{S}}_{\Sigma}$ of acceptable arguments of the argumentation system $\langle \mathcal{A}(\Sigma), \mathit{Undercut}, \mathit{Pref} \rangle$ 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* $\langle \mathcal{A}(\Sigma), \mathit{Undercut}, \mathit{Pref} \rangle$ *is:*

$$\underline{\mathcal{S}}_{\Sigma} = \bigcup \mathcal{F}^{i \geq 0}(\emptyset)$$

Note that since:

$$\mathcal{F}^0(\emptyset) = C_{Undercut,Pref}$$

it follows that:

$$\underline{\mathcal{S}}_{\Sigma} = C_{Undercut,Pref} \cup \left[\bigcup \mathcal{F}^{i \geq 1}(C_{Undercut,Pref}) \right]$$

An argument is *acceptable* if it is a member of the acceptable set. If the argument (H,h) is acceptable, we talk of there being an acceptable argument for h, and we say that the proposition h is *acceptable to* an agent that has an acceptable argument for it. An acceptable argument is one which is, in some sense, proven since all the arguments which might undermine it are themselves undermined.

It should be stressed that our choice of this particular argumentation system as a basis for the models of dialogue discussed below was somewhat arbitrary (it was the system that the third author studied in her Ph.D. thesis, and so is one that we are familiar with). What is important is that it has a notion of what an argument is, a notion of the strength of an argument, and a notion of an argument being acceptable—these are features that control the exchange of locutions in a dialogue. Any other argumentation system that has such features could be used as the basis of the dialogue systems discussed here without the need to change any of the technical details, albeit with the side effect of possibly changing some of the properties of the systems.

3 Locutions

As in our previous work [3, 6], 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. As discussed in [4] this gives argumentation-based dialogues a social semantics in the sense of Singh [41]—when agents assert something, they are committed to back up that something by giving the argument for it. 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, which we will call P and C[3]. Each agent has a knowledge base, Σ_P and Σ_C respectively, containing their beliefs. In addition, following Hamblin [17] 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 [6] for example) an agent's commitment store is just a subset of its knowledge base. Note that the union of the commitment 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

$$\langle \mathcal{A}(\Sigma_C \cup CS(P), \mathit{Undercut}, \mathit{Pref} \rangle$$

We denote the set of all arguments $\mathcal{A}(\Sigma_P \cup CS(C))$ by $\mathcal{A}(P, C)$.

All the knowledge bases contain propositional formulas and are not closed under deduction, and all are stratified by degrees of belief as discussed above. Here we assume that these degrees of belief are static and that both the players agree on them. As with the model of preferences itself, this is a very restrictive assumption, but once again we will stick with it for ease of explication. Elsewhere [5] we have discussed how 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 [27].

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, dialogue rules, and update rules. These are based on the rules suggested by [23] which, in turn, were based on those in the dialogue game DC introduced by MacKenzie [22]. The rationality rules specify the preconditions for making the move. Unlike those in [3, 6] 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 addresses move i of the dialogue to player C. The first move of the dialogue is move 1, $CS_0(P) = \emptyset$, $CS_0(C) = \emptyset$, and P and C strictly alternate. (A more formal description of the dialogue may be found in [5].) We start with the assertion of facts:

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

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

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 characters $\mathcal{U}$ and $\mathcal{PA}$, 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** $CS_i(P) = CS_{i-1}(P) \cup S$ 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.

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. This locution could easily be named "explain", but we inherit "challenge" from DC, and keep the name to make the heritage clear.

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 [3]—the main difference from the latter is that there are no "dialogue conditions" to specify the protocol. 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 [13, 34] and, unlike in some dialogue systems, there is no *retract* locution. Note that these locutions are ones used *within* dialogues—further locutions such as those discussed in [25] would be required for agents to agree to engage in dialogues, and to agree to switch between different kinds of dialogue.

4 Agent attitudes

One of the main aims of this paper is to explore how the kinds of dialogue in which agents engage depend 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. We deal first with assertion conditions. Note that we now name our agents G and H, to make it clear that either G or H can be the P or C of the previous section.

Definition 7 *An agent may have one of three* assertion *attitudes. If agent G is engaged in a dialogue with agent H, then:*

- *If G is* confident*, then it can assert any proposition p for which there is an argument $(S, p) \in \mathcal{A}(G, H)$.*
- *If G is* careful *then it can assert any proposition p for which there is an argument (S, p) if no stronger argument $(S', \neg p)$ exists in $\mathcal{A}(G, H)$.*
- *If G is* thoughtful *then it can assert any proposition p for which there is an acceptable argument $(S, p) \in \mathcal{A}(G, H)$.*

Thus a thoughtful agent will only put forward propositions which, so far as it knows, are correct. A careful agent will only put forward propositions which aren't directly rebutted. A confident agent won't stop to make either of these checks. Thus thoughtful and careful agents might be considered more discriminating informants than their confident counterparts, but neither can be considered more discriminating than the other:

Proposition 1 *Consider an agent G. If G is thoughtful or careful, then the assertions it can make are a subset of those that it could make were it confident.*

If G is thoughtful, then the set of assertions it can make overlap with those it could make were it careful.

Proof: If G is confident, it can assert any p for which it has an argument (S, p) that is in $\mathcal{A}(G, H)$. If G is careful, it can only assert those ps for which it has an argument (S, p) and no stronger argument for $\neg p$. These are clearly a subset of those for which it has an argument. If G is thoughtful, it can only assert p if (S, p) is in the set $\underline{\mathcal{S}}_{\Sigma_G \cup \Sigma_H}$. By Definition 6, $\underline{\mathcal{S}}_{\Sigma_G \cup \Sigma_H}$ can include arguments for propositions p for which there are stronger arguments for $\neg p$ (just so long as there are also even stronger arguments which undercut the arguments for $\neg p$), and so the first part of the result holds.

If there is an argument for p which is both acceptable and stronger than any argument for $\neg p$, then both a thoughtful G and a careful G could assert p. However, it is also possible that the argument for p could be stronger than any argument for $\neg p$ and also be undercut by a stronger argument for some element of its support which itself is not undercut. In this case a careful G could assert p but a thoughtful G could not. Finally, it is possible that there is an argument for $\neg p$ which is stronger than that for p and is undercut by a yet stronger argument which is not undercut. Then a thoughtful G could assert p and a careful G could not. $\square$

Given the fact that the set of possible assertions increases from thoughtful and careful to confident, and that, as we shall see, the ease with which these assertions can be computed increases also, 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. However, it is easy to show that:

Proposition 2 *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.

$$\langle \mathcal{A}(\Sigma), \mathit{Undercut}, \mathit{Pref} \rangle$$

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 inferred 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. $\square$

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. If agent G is engaged in a dialogue with agent H, then:*

- *If G is* credulous *then it can accept any proposition p previously asserted by H if $(S, p) \in \mathcal{A}(G, H)$.*
- *If G is* cautious *then it can accept any proposition $p$previously asserted by H for which there is an argument $(S, p) \in \mathcal{A}(G, H)$ if no stronger argument $(S', \neg p)$ exists in $\mathcal{A}(G, H)$.*
- *If G is* skeptical *then it can accept any proposition p previously asserted by H for which there is an acceptable argument $(S, p) \in \mathcal{A}(G, H)$.*

Again we can identify the relationship between the sets of propositions acceptable to different types of agent.

Proposition 3 *Consider an agent G. If G is skeptical or cautious, then the assertions it can accept are a subset of those it could accept were it credulous. If G is skeptical, then the set of assertions it can accept overlaps with the set of assertions it could accept were it cautious.*

Proof: Consider an agent G such that $(S, p) \in \mathcal{A}(G, H)$. By definition, if G is credulous, it can accept any q for which it is presented with an argument (S', q). It can therefore accept $\neg p$ provided it is given $(S', \neg p)$ even if $(S, p) \gg^{Pref} (S', \neg p)$. If G were cautious, then it would not be able to accept $\neg p$ unless $(S', \neg p) \gg^{Pref} (S, p)$. Thus a cautious agent can only accept a subset of the arguments that a credulous agent can accept. If G is skeptical, it might not accept $\neg p$ even if $(S', \neg p) \gg^{Pref} (S, p)$, because another argument (S'', r) might exist which undercuts $(S', \neg p)$ and makes $(S', \neg p)$ unacceptable. Thus a skeptical agent can only accept a subset of the arguments that a credulous agent can accept, and the first part of the result is proved.

If it is presented with an argument for p which is both acceptable given what it knows and is and stronger than any argument it has for $\neg p$, then both a skeptical G and a cautious G could accept p. However, it is also possible that the argument for p could be stronger than any argument G has for $\neg p$ and also be undercut by a stronger argument which G has for some element of its support which itself is not undercut. In this case a cautious G could accept p but a skeptical G could not. Finally, it is possible that G has an argument for $\neg p$ which is stronger than that for p and is undercut by a yet stronger argument which is not undercut. Then a skeptical G could accept p and a cautious G could not. $\square$

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 *querulous* 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 4 *The set of propositions acceptable to a skeptical agent using an argumentation system $\langle \mathcal{A}(\Sigma), \mathit{Undercut}, \mathit{Pref} \rangle$ is exactly the same as the set of propositions acceptable to a querulous agent using the same argumentation system.*

Proof: Consider a skeptical agent G and a querulous 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. $\square$

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 querulous agent adds nothing to our classification.

Since agents will typically both assert and accept propositions during a dialogue, both their assertion attitudes and their acceptance attitudes need to be specified. We write $\langle \mathit{accept} \rangle / \langle \mathit{assert} \rangle$ to denote an agent with acceptance attitude $\langle \mathit{accept} \rangle$ and assertion attitude $\langle \mathit{assert} \rangle$. With a pair of agents that are skeptical/thoughtful, we recover the rationality conditions of the dialogue system in [3].

5 Dialogue types

With the agent attitudes specified, we can begin to look at different types of dialogue in detail giving protocols for each. Note that these are very simple protocols, intentionally so. Indeed they are the simplest protocols we can think of that meet the criteria for the different types of dialogue laid down by Walton and Krabbe. As a result these protocols are very rigid, and more flexible protocols will very likely be required. However, what we aim to do here is to establish a baseline by looking at the properties of these simple protocols before going on to examine the properties of more complex protocols (such as those we have defined in [24]).

An important feature common to all these protocols is that if an agent repeats a locution, then the dialogue terminates. We call protocols with this feature *non-circular* protocols.

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 (again the name change distinguishes these agents, which play particular roles in a dialogue, from G and H, which can take any role), 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 $assert(p)$ or $assert(\neg p)$ if it can, and $assert(\mathcal{U})$ if it cannot. Which response is given will depend upon the contents of its knowledge-base and its assertion attitude[5]. $\mathcal{U}$ indicates that, for whatever reason B cannot give an answer.
3. A either *accepts* B's response, if its acceptance attitude allows, or *challenges*. $\mathcal{U}$ cannot be *challenged* 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.
6. A accepts p if its acceptance attitude allows.

Note that A accepts whenever possible, and is only 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. B then has to give a response if it has one. All of these seem to us to be reasonable conditions. More flexible dialogue protocols are allowed, as in [3], but at the cost of possibly running forever[6].

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 5 *When subject to challenge(p) for any p it has asserted, a confident, careful, 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, confident, careful 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.

[5] It might even be able to assert both p and $\neg p$, in which case it chooses one and asserts that.

[6] The protocol in [3] allows an agent to interject with $question(p)$ for any p at virtually any point, allowing two agents to prolong a dialogue indefinitely by issuing endless *questions* about arbitrary formulae.

For a proposition to be in S by Definition 1 it must be part of a consistent, minimal subset of $\Sigma_G \cup CS(H)$ (where H is the other agent in the dialogue) which entails p. Any such proposition q is the conclusion of an argument $(\{q\}, q)$ and this argument is easily generated. $\Box$

This first result ensures that step 4 can always follow from step 3, and the dialogue will not get stuck at that point. 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 proposition asserted by an agent:

Proposition 6 *A credulous agent G operating under protocol $\mathcal{IS}$ will always accept a proposition asserted by a confident, careful 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 5, 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. $\Box$

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

Proposition 7 *All dialogues carried out using non-circular protocols and the set of moves $\mathcal{M}'_{\mathcal{DC}}$ will terminate.*

Proof: We have agents with finite knowledge-bases, a set of locutions which are instantiated with some subset of the knowledge-bases, and protocols that terminate the dialogue if an agent repeats itself. If the dialogue does not end before every possible locution is made, then it will end once the (finite) set of possible locutions have all been made once. $\Box$

This, of course, does not bound the length of the dialogue very tightly. Since agents are allowed to assert sets of propositions, it is conceivable that an agent G can make $O(|2^{\Sigma_G}|)$ moves before it repeats itself, so a dialogue between G and another agent H might take as many as $O(|2^{\Sigma_G \cup \Sigma_H}|)$ steps (since as soon as one agent asserts something the other can use it in an argument). We can get much tighter bounds on the length of the dialogue by considering the protocol in detail:

Proposition 8 *An information-seeking dialogue under protocol $\mathcal{IS}$ between a credulous, cautious or skeptical agent G and a confident, careful or thoughtful agent H, where H moves first, will always terminate in $O(|\Sigma_H|)$ moves.*

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 credulous, then by Proposition 6, 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 a stronger argument for $\neg p$. In the former case the dialogue terminates immediately. In the latter case G will challenge p and by Proposition 5 receive the support S. If G doesn't have an argument against any of the $s \in S$, then they will be accepted, and this may be enough to make G accept p in which case the dialogue will terminate. If not, the only locution that G could utter is *challenge*(p), 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 5 this will produce an argument $(\{s\}, s)$ from H, and G will not be able to accept this any more than it could accept the s initially. G will therefore challenge s, which repeats its previous locution and so 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 6), 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 will therefore repeat its last challenge, and the dialogue will terminate. Since the behaviour of H only depends on it having an argument for p or $\neg p$, the result holds whether H is confident, careful or thoughtful.

In the worst possible case, the dialogue will run on until G has examined all of the s in turn (and either accepted all of them or accepted all but one, challenged this last, found it unacceptable and then terminated the dialogue unsuccessfully with another challenge), and in the very worst case the S in question will be the whole of Σ_H. $\square$

This result gives us much tighter bounds on the number of steps than Proposition 7, and, of course, on average the number of steps will be even less since the set of propositions in an argument will usually be much less than the whole knowledge-base of an agent.

While this result is a good one, because of the guarantee of quick termination, the proof illustrates a limitation of the dialogue protocol and the way that the agents handle utterances in the dialogue. If G is skeptical or cautious, it may never come to accept p whatever H says. That is H may not persuade G to change its mind even though it has information which undermines G's argument for $\neg p$. 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 should engage in a persuasion dialogue, albeit one that is embedded in an information seeking dialogue as in [25]. However, since persuasion dialogues suffer from similar problems, we return to this limitation in Section 5.4.

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 [24]). 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 [25], and from this point can adopt the following protocol $\mathcal{I}$:

1. A asserts $q \to p$ for some q or $\mathcal{U}$.

2. B accepts $q \to p$ 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 $q \to p$ by s.

5. B accepts $q \to p$ if its acceptance attitude allows, or the dialogue terminates.

6. B *asserts* q, or $r \to q$ for some r, or $\mathcal{U}$.

7. A accepts the previous assertion if its acceptance attitude allows, or challenges it.

8. B replies to a *challenge* with an *assert*(S), where S is the support of an argument for the last proposition challenged by B.

9. Goto 7 for each proposition $s \in S$ in turn.

10. A accepts B's assertion in 6 if its acceptance attitude allows, or the dialogue terminates.

11. If $\mathcal{A}(CS(A) \cup CS(B))$ includes an argument for p which both agents' attitude allows them to accept, then the dialogue terminates.

12. Go to 1, substituting r for p and some t for q.

Note also that in step 2, when agent B makes an assertion $q \to p$ then it is bound to not know q (so its assertion is almost counterfactual)—if it did know q, then the initial conditions for the inquiry would not be met. However, for all subsequent steps in the dialogue, the agent making the assertion $r \to t$ might also know r. Note also that the protocol could equally well have been written with agent B making step 1—the decision to make A the first agent to move was arbitrary unlike the case for the information seeking dialogue.

This protocol is basically a series of implied $\mathcal{IS}$ dialogues. First B asks "do you know of anything which would imply p were it known?". A replies with one, or the dialogue terminates with $\mathcal{U}$. If B does not accept the implication as for an information seeking dialogue, the dialogue terminates unsuccessfully. If B does accept the implication, A 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 and which completes the chain of implications, or there is no new implication to add to the chain. Because of this structure, it is easy to show that:

Proposition 9 *An inquiry dialogue $\mathcal{I}$ between two agents G and H with any acceptance and assertion attitudes will always terminate in $O(|\Sigma_G \cup \Sigma_H|)$ steps.*

Proof: The dialogue starts with what is effectively an implied $\mathcal{IS}$ dialogue and runs exactly as in Proposition 8. If it terminates successfully (that is with a result other than $\mathcal{U}$ or B not accepting A's assertion), then it is followed with a second $\mathcal{IS}$ dialogue in which the roles of the agents are reversed. Again this dialogue will runs exactly as in Proposition 8, possibly ending 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 first. 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.

Now, from Proposition 8, we know what the worst case length of each of these iterated dialogues is. Since each can in theory involve $O(|\Sigma_G|)$ or $O(|\Sigma_H|)$ steps, it might appear that this dialogue can run for much longer than one under $\mathcal{IS}$, running $O(|\Sigma_G|)$ or $O(|\Sigma_H|)$ for each implication in the proof. However each of the propositions in Σ_G and Σ_H can only be asserted once at most so, no matter how many sub-dialogues there are, there can be at most $O(|\Sigma_G \cup \Sigma_H|)$ steps. □

This simple protocol, like that for information seeking, is flawed, and this time we will consider ways to fix the flaws. One problem is that $\mathcal{I}$ may not permit a proof to be found even though one is available to the agents if they were to make a different set of assertions. More precisely, we have:

Proposition 10 *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 $\mathcal{A}(\Sigma_G \cup \Sigma_H)$ provides an argument p which both agents would be able to accept.*

Proof: Assume 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. □

There is another flaw in the structure of the dialogue. As it stands it assumes that agents can take strict turns in constructing the proof. If an agent cannot fill in a new step in its turn, its only alternative is to to utter $\mathcal{U}$ and bring the dialogue to an end. More formally this means that there is another kind of case that could prove Proposition 10, namely that in which G has $\Sigma_G = \{q \rightarrow p\}$ and H has $\Sigma_H = \{r \rightarrow q, r\}$. In our experience of inquiry dialogues, albeit ones between human agents, it is not unusual for one agent to fill in several steps—indeed it is much more common than for agents to strictly alternate.

Of course, it is possible to design protocols which don't suffer from these problems, for example 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) and by allowing agents to explicitly "pass" if they cannot add to the proof. For example we have the protocol $\mathcal{I}'$:

1. A asserts either $\bigcup_i \{q_i \rightarrow p\}$ for all $q_i \rightarrow p$ which its assertion attitude allows it to assert or $\mathcal{PA}$.
2. If A asserts $\mathcal{PA}$, then goto 1 switching agent roles.
3. For each i in turn, B accepts $q_i \rightarrow p$ if its acceptance attitude allows, or challenges it.
4. A replies to a *challenge* with an *assert*(S), where S is the support of an argument for the last proposition challenged by B.
5. Goto 3 for each proposition $s \in S$ in turn, replacing $q_i \rightarrow p$ by s.
6. B accepts any of the challenged $q_i \rightarrow p$ if its acceptance attitude allows.
7. If B has accepted none of the $q_i \rightarrow p$ then the dialogue terminates.
8. B *assert*s either some q_i, or $\bigcup_j \{r_j \rightarrow q_i\}$ for all $r_j \rightarrow q_i$ that its assertion attitude allows it to assert or $\mathcal{PA}$.
9. If B asserts $\mathcal{PA}$, then A asserts $\mathcal{PA}$, q_i or $r_j \rightarrow q_i$ for $q_i \rightarrow p$. Goto 3 replacing $q_i \rightarrow p$ with whatever A asserted.
10. For each i in turn, A accepts $r_j \rightarrow q_i$ if its acceptance attitude allows, or challenges it.
11. B replies to a *challenge* with an *assert*(S), where S is the support of an argument for the last proposition challenged by A.
12. Goto 10 for each proposition $s \in S$ in turn, replacing $r_j \rightarrow q_i$ by s.
13. A accepts any of B's assertions in 10 that it challenged if its acceptance attitude allows.
14. If A has accepted none of B's assertions in 10, then the dialogue terminates.
15. If $\mathcal{A}(CS(A) \cup CS(B))$ includes an argument for p which both agents can accept, then the dialogue terminates successfully.

16. Go to 1, substituting r_k for p and t_l for q_i.

The locution $\mathcal{PA}$ indicates that the agent "passes", and if two are uttered in sequence then the dialogue terminates. We can easily show that:

Proposition 11 *An inquiry dialogue $\mathcal{I}'$ between two agents G and H with any acceptance and assertion attitudes will always terminate in $O(|\Sigma_G \cup \Sigma_H|)$ steps.*

Proof: An inquiry dialogue under $\mathcal{I}$ starts like an information seeking dialogue in which the question has already been asked. An inquiry dialogue under $\mathcal{I}'$ starts like a set of information seeking dialogues all with the same initial question since A can reply with a set of answers, and these will run as in Proposition 8. If all terminate without B accepting any of A's assertions, then the dialogue ends unsuccessfully.

If there is no $\mathcal{PA}$, and one of the sub-dialogues terminates successfully, then it is followed with a second $\mathcal{IS}$ dialogue in which the roles of the agents are reversed and B is expected to start a new set of sub-dialogues for every implication that it accepted. Again these will run as in Proposition 8, and the sub-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 $\mathcal{PA}$, then it is followed with another set of $\mathcal{IS}$ dialogues in which the roles of the agents are again reversed. This third dialogue set runs just like the first. The iteration will continue until one of the agents utters a $\mathcal{PA}$ (indicating it has nothing else it can legally say), one repeats itself, or the chain of implications is ended. One of these things 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).

Any time that a $\mathcal{PA}$ is uttered, the agent that did not utter it becomes the agent that must assert something. That will either be a $\mathcal{PA}$, or a new step in the proof (starting a new cycle of sub-dialogues, all of which will eventually end (as argued above) with either the completion of the proof, a repetition or a $\mathcal{PA}$. Thus we will either get two $\mathcal{PA}$s in a row, ending the dialogue unsuccessfully, the proof will be completed and we will have a successful termination, or the dialogue will end unsuccessfully with a repetition.

Thus it is clear that there is nothing in the protocol $\mathcal{I}'$ that significantly increases the number of steps in the worst case with respect to $\mathcal{I}$. There will be more steps typically, because more assertions of implications will be made, and so there will be more challenges and assertions of grounds. However, the agents still terminate the dialogue if they assert the same proposition twice, and so cannot increase the number of steps above $O(|\Sigma_G \cup \Sigma_H|)$. $\square$

Although this protocol will solve both the problems with $\mathcal{I}$ outlined above, it won't ensure that a proof is found if one exists because only the agent which currently "holds the floor" is allowed to make assertions, these are restricted to those assertions which connect to things that have just been uttered, and an agent will only $\mathcal{PA}$ if it has nothing to say. Thus a critical step in the proof might be passed over if the agent that knows it is not able to utter it at the right place because the other agent is saying something which, although it connects to the proof tree the agents are jointly constructing, it is not on a path that ultimately leads to a proof. One might, of course, further improve the protocol by allowing agents to assert anything which extends the proof tree at any point, but doing this would lead us a bit too far from our aim, which is to look at some simple protocols which allow agents to carry out various kinds of dialogue—we now have two inquiry dialogue protocols and any further dialogue would be rather more complex than we have set out to define here. It is time to move on.

It is worth noting that, as hinted at above, in contrast to the information seeking dialogue, in any inquiry dialogue 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, for instance, 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 [3] are based, is fundamentally a persuasion game, so the protocol below results in games which are very like those described in [3]. 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.
5. B accepts p if its acceptance attitude allows, or the dialogue terminates.

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

Once again the form of this dialogue has much in common with information seeking 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 $\mathcal{IS}$ 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 12 *A persuasion dialogue under protocol $\mathcal{P}$ between two agents G and H will always terminate in $\max(O(|\Sigma_G|), O(|\Sigma_H|))$ steps.*

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 how a dialogue under $\mathcal{IS}$ proceeds, it suffices to consider how the assertion of $\neg p$ affects things. Since the only difference between the sub-dialogue spawned by the assertion of $\neg p$ and an $\mathcal{IS}$ 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 the dialogue can proceed differently. However, this assertion of p repeats the assertion that provoked the $\neg p$ and so the dialogue terminates immediately.

Thus a $\mathcal{P}$ dialogue in which G moves first will either have H challenging and then end after the examination of the grounds for p by H in a maximum of $O(|\Sigma_G|)$ steps, or will have H assert $\neg p$ and then terminate after G has examined the grounds for $\neg p$ in a maximum of $O(|\Sigma_H|)$ steps. Thus the dialogue will terminate in at most $\max(O(|\Sigma_G|), O(|\Sigma_H|))$ steps. $\square$

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 many persuasion dialogues in the literature.

5.4 Limitations of the protocols

As mentioned above, the protocols we have discussed here are intentionally simple. As a result, while the protocols capture the essential features of the types of dialogue as defined by Walton and Krabbe [45], they have a number of limitations.

The main limitation is the behaviour of the set of steps common to all the dialogues. One agent asserts a proposition p, the other either immediately accepts it, or challenges it and is then faced with a new set of propositions S. The $s \in S$ are then individually accepted or challenged, and the dialogue terminates at the end of this exchange with either p and S accepted or without p and at least one s having been accepted (as soon as one s is rejected the dialogue must end unsuccessfully).

While, as argued above, this is perfectly reasonable for information seeking dialogues, it means that only a very limited form of persuasion is possible. For A to persuade B to accept p when B initially accepts $\neg p$, the support for p that A asserts must all be facts that are higher in the preference order than any arguments to the contrary that B may possess (since the argument inherits its strength from its weakest link). If B has any stronger counter-argument, the dialogue ends without A being able to engage B in a persuasion about the grounds for B's counter-argument. This forced termination seems rather unnatural. Much more natural would be to allow persuasion to continue in the vein sketched above, with agents permitted to engage in this kind of counter-argumentation and even to backtrack to examine propositions that were asserted several moves earlier in the dialogue. This kind of backtracking would extend the flexibility of inquiry dialogues as well, and is one of the things we hope to look at in the future.

This is also the place to note that, of course, the kind of persuasion we are dealing with here is one that relies on the supply of new information. It allows us to capture the following kind of dialogue (though to capture exactly this dialogue would require a slightly different protocol):

A: I believe that Henry Kissinger is a bad man.

B: I believe that Henry Kissinger is not a bad man.

A: There is evidence that he helped to prolong the Vietnam war and so caused unnecessary deaths among American soldiers and the Vietnamese population in general (to say nothing of the bombing of Cambodia).

B: What is your evidence for this assertion?

A: It is contained in Christopher Hitchens' book "The Trial of Henry Kissinger" [18].

B: (after examining the evidence in [18]). I did not know those things. They outweigh my arguments for Kissinger not being a bad man. I now believe that Henry Kissinger is a bad man.

However we cannot handle a similar kind of dialogue in which B initially did know these facts, thought that an argument based on Kissinger's role as a statesman was stronger than the argument concerning prolonging the war, but was persuaded to reverse this preference (maybe by an appeal to what B would think if one of the aforementioned deaths was their child). Persuasion dialogues of this latter kind, however, could be handled by using the dialogue system defined in [5].

6 Properties of agent attitudes

In this section, we consider the result of the dialogues, in particular with respect to the arguments the agents end up accepting and how these relate to the agents' attitudes. The following results hold for all dialogue types, but, as we will see, some are more applicable to different types of dialogue, so it is helpful to have a formal means of distinguishing the different types. To do this we need some additional definitions:

Definition 9 *An agent is said to* entertain *a proposition p if it has an argument A_1 for p.*

Definition 10 *An agent is said to* believe *a proposition p if it has an argument A_1 for p and this is stronger than any argument A_2 for $\neg p$.*

Thus we use the term "believe" in the sense of "believed more strongly than the contrary by direct proof". We also refer to p being *more strongly believed* than q if argument A_1 for p is stronger than any argument for q.

Note that both this idea of belief and the notion of acceptability are not subject to the law of the excluded middle in the sense that it is quite possible for an agent to have neither an acceptable argument for p nor for $\neg p$ and to believe neither p nor $\neg p$. As we will see below, it is also possible for an agent to believe p but not have an acceptable argument for it, and for p to be acceptable to an agent which does not believe it. In part we make these distinctions because it is computationally simpler (exactly how much simpler will be discussed in Section 7) for agents to identify entertained propositions than believed propositions, and to identify believed propositions than acceptable ones. Thus there may be computational advantages to building confident agents over thoughtful ones and credulous agents over skeptical ones.

We also distinguish:

Definition 11 *An agent is said to* be sure of *a proposition p if it has an argument A_1 for p and this is stronger than any undercutting argument A_2.*

These terms allow us to distinguish the different kinds of dialogue. An information seeking dialogue about p opens with A entertaining neither p nor $\neg p$ (and thus not believing or having an acceptable argument for either), while B may entertain p, $\neg p$, both or neither (though A believes that B at least entertains p or $\neg p$). An inquiry opens with neither A nor B entertaining either p or $\neg p$. In contrast, a persuasion dialogue opens with one agent having an argument for p that accords to its own acceptance criterion, and the other either not having an argument for p that accords to its own acceptance criterion, or having an argument for $\neg p$ that accords to its own acceptance criterion.

Clearly these notions are related to each other and also to the notion of acceptability:

Proposition 13 *Consider an agent G with an argumentation system AS_G.*

1. *Any proposition that is believed by G is entertained by G.*

2. *A proposition that is entertained by G is not necessarily believed by G.*

3. *Any proposition that G is sure of is also believed by G.*

4. *A proposition that is believed by G is not necessarily one that G is sure of.*

5. *A proposition that is acceptable to G is not necessarily believed by G.*

6. *A proposition that is believed by G is not necessarily acceptable to G.*

7. *Any proposition that G is sure of is also acceptable to G.*

8. *A proposition that is acceptable to G is not necessarily one that G is sure of.*

Proof: These properties follow almost immediately from the definitions and the fact that an argument can be both believed and undercut:

1. For an agent to believe a proposition p, it has to have a stronger argument for it than $\neg p$. The agent therefore has an argument for p, and so entertains it.

2. An agent which has an argument for a proposition p and a stronger argument for $\neg p$ will entertain p without believing it.

3. For an agent to be sure of a proposition p it has to have a stronger argument for p than any undercutting argument. An argument for $\neg p$ is an undercutter, so must be weaker than the argument for p. Thus p is believed.

4. For an agent to believe a proposition p, it has to have a stronger argument for it than its negation. However, it can have an undercutter which is stronger than the argument for p, and thus the agent is not sure of p.

5. For a proposition p to be acceptable to an agent, the agent has to have a stronger undercutter for any argument which undercuts the argument for p. Thus p will be acceptable if there is an argument for p, a stronger argument for $\neg p$, and an even stronger argument which undercuts the argument for $\neg p$. In this latter case the agent will not believe p.

6. For an agent to believe a proposition p, it has to have a stronger argument for p than $\neg p$. However, the agent can have an undercutter which is stronger than the argument for p, and, if this undercutter is itself not undercut by a stronger argument, this will prevent p from being acceptable.

7. For an agent to be sure of a proposition p it has to have a stronger argument for p than any undercutting argument. This means that the argument defends itself against any undercutter and so is acceptable, making p acceptable to the agent.

8. For a proposition p to be acceptable to an agent, the agent has to have a stronger undercutter for any argument which undercuts the argument for p. Thus it p will be acceptable if there is an argument for p, a stronger argument which undercuts the first, and an even stronger argument which undercuts the second. In this latter case the agent will not be sure of p.

Thus the Proposition follows. □

Propositions that an agent is sure of are a special case of acceptable propositions—they are ones which only have to be checked for undercutters. There is no need to look for undercutters of the undercutters (and so on) since the original arguments defend themselves. Propositions that an agent is sure of are also a special case of propositions that are believed—in essence they are propositions for which every element of their support is believed. We could therefore define a new "super-thoughtful/super-skeptical" class of agent which only asserts and accepts propositions which it is sure of, and we would find that it asserts and accepts propositions which are in the intersection between those which can be asserted and accepted by, respectively, thoughtful and careful agents and skeptical and cautious agents.

The classification of propositions can also be related to what agents can assert, a result which ties in with Proposition 1:

Proposition 14 *A confident agent can assert any proposition it entertains, believes, is sure of, or which is acceptable to it. A careful agent can assert any proposition which it believes, or which it is sure of. A thoughtful agent can assert any proposition which it is sure of or which is acceptable to it.*

Proof: Immediate from the definitions of the types of proposition, the definition of agent attitudes, and Proposition 13. □

Underlying information seeking dialogues is the idea that agents have a reasonably benevolent attitude to one another. Partly this is implicit in the fact that A is requesting information from B—if B is unhelpful it need never answer. More importantly, perhaps, is the fact that B might be able to mislead A. Of course if B grounds its reply in facts that A knows to be untrue then A may not accept the reply, but obviously this depends upon its acceptance attitude. As it turns out, there are a number of situations in which A and B can engage in an information seeking dialogue which results in A accepting an argument while one or other of them has a stronger argument for the opposite. These dialogues can be considered pathological for this reason and so warrant further study. To investigate such situations we need the following:

Definition 12 *An agent G is said to* convince *an agent H about the truth of a proposition p if G asserts p and H accepts it.*

We say that one agent *misleads* another if it manages to convince the second of something the first would not accept itself, and we now look to see under what circumstances one agent can mislead another, easily obtaining some simple results which clarify the situation. Note that here, as throughout the paper, we assume that an assertion p is always immediately followed by an assertion of its support.

Proposition 15 *An agent G can convince an agent H of p even if H does not believe p.*

Proof: Whatever the assertion type of G it must have an argument (S, p) in order to be able to assert p. If H is credulous, then by definition it will accept p and thus be convinced. If H is cautious, then it will be convinced if it has no argument $(S', \neg p)$ such that $(S', \neg p) \gg^{Pref} (S, p)$. Similarly, if H is skeptical, it will still be convinced of p unless it has some argument (S'', q) such that (S'', q) undercuts (S, p). $\square$

This result, which is very weak in the sense that it only shows the possibility of H being convinced rather than that H will be convinced, holds no matter what acceptance attitude of H. If H is credulous, then we can get a much stronger result:

Proposition 16 *An agent G will always convince a credulous agent H of p even if H believes $\neg p$.*

Proof: Whatever the assertion type of G it must have an argument (S, p) in order to be able to assert p. If H is credulous, then by definition it will accept p and thus be convinced, even if it itself has $(S', \neg p)$ such that $(S', \neg p) \gg^{Pref}$ (S, p). $\square$

Thus, as one might expect, allowing an agent to be credulous means that it is open to exploitation no matter how well informed it is. However, if H is cautious or skeptical, then, as the proof of Proposition 15 makes clear, it will only be misled in this way if it has no stronger information to the contrary. In other words:

Proposition 17 *An agent G can only convince a cautious agent H of p if H does not believe $\neg p$ more strongly than G, and can only convince a skeptical agent H of p if $\mathcal{A}(\Sigma_G) \cup \mathcal{A}(\Sigma_H)$ contains an acceptable argument for p.*

Proof: Suppose G asserts (S, p). By definition, a cautious H can only accept p if it has no argument $(S', \neg p)$ such that $(S', \neg p) \gg^{Pref} (S, p)$. Thus a cautious H can only accept p if it does not believe $\neg p$. A skeptical H will only accept p if it has an acceptable argument for it, or one can be constructed once G has asserted sufficient information to undercut any undercutters H might have for (S, p), which will only happen if (S, p) is in the set of acceptable arguments of $\mathcal{A}(\Sigma_G) \cup \mathcal{A}(\Sigma_H)$. $\square$

Note that there is a certain asymmetry in this result which follows from the different kinds of check that the different attitudes require and that fact that a proposition can be believed but not acceptable to a given agent, and can be acceptable but not believed. In fact we can easily see that:

Proposition 18 *An agent G can convince a skeptical agent H of p even if H believes $\neg p$, and can convince a cautious agent H of p even if $\mathcal{A}(\Sigma_G) \cup \mathcal{A}(\Sigma_H)$ does not contain an acceptable argument for p.*

Proof: Both of these possibilities follow directly from the acceptance attitudes of the types of agent and the results in Proposition 13. □

Thus, taking the obvious corollary, even agents which try to ensure the quality of information they are given can be misled when they don't have the right information with which to check what they are told or don't apply this information in the right way (because of their acceptance attitude). This is particularly important in information seeking and inquiry dialogues since, by definition, in such dialogues the agents which initiate the dialogue do not have this information. Indeed, in information seeking dialogues under $\mathcal{I}$ or $\mathcal{I}'$, Propositions 15–18 show that the dialogue may terminate with the agent which initiated the dialogue being convinced of something which it would not assert itself.

This, of course, does not amount to misleading as we have defined it since the definition includes the proviso that the agent which is asserting the proposition in question would not accept it itself, so we have to look a bit further. Consider a dialogue between agents A and B. Without loss of generality, we can consider that A is trying to persuade B that p is true, and so the critical attitudes here are the assertion and acceptance attitudes of A and the acceptance attitude of B. The three acceptance attitudes and three assertion attitudes would seem to give us a space of 27 combinations of agent attitude to consider. However, not all combinations of acceptance and assertion attitude of A are sensible since some will allow agents to assert things which they would not accept themselves. We define:

Definition 13 *An agent is* reliable *if it is only able to assert propositions which it would always accept itself.*

Then for an agent to be misleading, it has to not be reliable and be able to convince another agent to accept a proposition which it wouldn't accept itself. We have:

Proposition 19 *An agent G that is cautious/confident, cautious/thoughtful, skeptical/careful or skeptical/confident is not reliable. All other agents are reliable.*

Proof: A confident agent can assert any proposition p for which it has an argument (S, p). It may well have a stronger argument $(S', \neg p)$ as well. If it is cautious or skeptical it would therefore not accept p and so is not reliable. If it is credulous it will accept any argument and thus be reliable.

If the agent is careful it will not only have an argument for p (which will be acceptable to its credulous self), but it will also believe p so the argument will be acceptable to its cautious self. However, there might be a stronger undercutter for p (if p is not a proposition it is sure of), which might make the agent unable to get its skeptical self to accept p.

If the agent is thoughtful it can only assert p if it has an an acceptable argument. This would be accepted by a skeptical agent (by definition) and by a credulous agent since it is backed by an argument. A cautious agent however might not accept it since the argument (S, p) might be undercut by a stronger argument for $\neg p$ (making p not believable) but then rehabilitated by a stronger undercutter of this second argument. This completes the proof. $\square$

Obviously a reliable agent cannot be misleading. Thus any agent that is credulous cannot be misleading because it will itself believe whatever it says, and a cautious/careful agent and a skeptical/thoughtful agent will not be misleading because it will perform the right kind of check on what it says (according to its acceptance attitude) before making an assertion. Ignoring such agents for the time being, we are ready to identify the conditions under which misleading will occur:

Proposition 20 *In a dialogue between agents G and H, G can only mislead H about the proposition p if:*

- *G is cautious/confident, cautious/thoughtful, skeptical/careful or skeptical/confident and H is credulous or cautious; or*
- *G is cautious/confident or cautious/thoughtful and H is skeptical and $\mathcal{A}(\Sigma_G) \cup \mathcal{A}(\Sigma_H)$ contains an acceptable argument for p.*

Proof: If H is credulous, it will accept anything it is told, so any set of assertion and acceptance attitudes that render G not reliable will allow G to state something it would not accept itself and get H to accept it. If H is cautious then by Proposition 18 it will accept p provided it doesn't have a stronger argument against it. So, it is possible that any G which can make assertions it won't accept can mislead H. Finally, if H is skeptical then by Proposition 18 it will only accept p if $\mathcal{A}(\Sigma_G) \cup \mathcal{A}(\Sigma_H)$ contains an acceptable argument for p. This means that G can't be skeptical otherwise it would be able to accept p as well (meaning that no misleading was going on). Thus misleading will only happen for skeptical H if G is cautious/confident or cautious/thoughtful. $\square$

Note that this result only tells us when misleading might occur, not when it will happen since to predict the latter we would need to look at exactly what was in the knowledge bases of the agents.

Having identified the types of agent which can mislead, and having thus provided a means of ruling out some kinds of dialogues which give results we might consider unreasonable, we turn to trying to obtain some guarantees about how dialogues might give reasonable results. To do this we will clearly disregard agent types that are not reliable. In addition we will discard credulous agents since they will always accept any argument that is put forward and are thus easy to mislead. As a result, the only combinations we will consider are agents which are cautious/careful and skeptical/thoughtful, and we will only consider dialogues between agents that are of the same type.

We can show that these pairs of agents end dialogues under circumstances that seem reasonable given their attitudes. We have:

Proposition 21 *In a dialogue between two cautious/careful agents G and H, G will only convince H of p if G believes p more strongly than either agent believes $\neg p$.*

Proof: For G to convince H of p, it has to assert p and have it accepted. To assert p, since G is careful, then from Proposition 13 it must believe p (being sure of a proposition implies it is believed) and so by Definition 10 believe it more strongly than $\neg p$. For H to accept p it must not have a stronger argument for $\neg p$ than G puts forward, and so must not believe $\neg p$ more strongly than G believes p. $\square$

and

Proposition 22 *In a dialogue between two skeptical/thoughtful agents G and H, G will only convince H of p if, after the dialogue, p is acceptable to both agents.*

Proof: The proof is trivial but worth stating because it sheds some light on the way that the dialogues proceed. For G to convince H of p, it has to assert p and have it accepted. To assert p, G must have an acceptable argument for p. For H to accept p, either it finds p acceptable immediately, or G is able to assert arguments which outweigh whatever arguments initially make p not acceptable, thus making it acceptable. $\square$

The key difference between these two results is that with cautious/careful agents it is computationally much simpler to determine if one agent will convince another—one just examines the arguments which directly relate to the proposition in question. With skeptical/thoughtful agents the determination is more complex, and to some extent hangs on exactly what the agents say at certain points in the dialogue (a subject investigated in [32]), which is the reason for the condition "after the dialogue" in Proposition 22.

These results shows that pairs of cautious/careful agents and skeptical/thoughtful agents have reasonable behaviour. In a dialogue between cautious/careful agents, no agent will be convinced of something unless both it and its opponent believe it. If we consider the relation of undercutting to indicate one argument throwing doubt on another, then by the obvious extension of this result in a dialogue between two skeptical/thoughtful agents either will only be convinced of p when neither agent has any reason to throw doubt on p.

Note that this does not ensure that the agents cannot be mistaken (they may just lack the information) and it does not prevent one agent lying to another, provided it has a suitably strong reason to want to tell the lie. The result can also be taken as validating the choice in [3] to make all agents skeptical/thoughtful, but also suggests that cautious/careful agents warrant further study.

7 Complexity Results

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 in order that we can reduce it by, for example as we did in [46], suitable choice of language.

We begin with a brief survey of the relevant key concepts from complexity theory (see, e.g., [26] for detailed definitions). We start with the complexity classes P (of languages/problems that may be recognised/solved in deterministic polynomial time), and NP (of languages/problems that may be recognised/solved in non-deterministic polynomial time). If $\mathcal{C}$ and $\mathcal{C}'$ are complexity classes, then we denote by $\mathcal{C}^{\mathcal{C}'}$ the class of languages/problems that are in $\mathcal{C}$ assuming the availability of an oracle for languages/problems in $\mathcal{C}'$ [26, pp415–417]. Thus, for example, $\mathrm{NP}^{\mathrm{NP}}$ denotes the class of languages/problems that may be recognised/solved in non-deterministic polynomial time, assuming the presence of an oracle for languages/problems in NP. A language that is complete for $\mathrm{NP}^{\mathrm{NP}}$ would thus be NP-complete even if we had "free" answers to NP-complete problems (such as propositional logic satisfiability). We define the *polynomial hierarchy* with reference to these concepts [26, pp423–429]. First, define

$$\Sigma_0^p = \Pi_0^p = \mathrm{P}$$

Thus both Σ_0^p and Π_0^p denote the classes of languages/problems that may be recognised/solved in deterministic polynomial time. We then inductively define the remaining tiers of the hierarchy, as follows:

$$\Sigma_{u+1}^p = \mathrm{NP}^{\Sigma_u^p} \qquad \Pi_{u+1}^p = \text{co-}\Sigma_{u+1}^p$$

Thus Σ_1^p is simply the class NP, and Π_1^p is the class co-NP, while $\Sigma_2^p = \mathrm{NP}^{\mathrm{NP}}$ and $\Pi_2^p = \text{co-}\mathrm{NP}^{\mathrm{NP}}$.

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 23 *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 24 *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 $v' \models H \to h$.

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 $H \to h$ 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 $\text{QBF}_{2,\exists}$ problem [19, p96]. An instance of $\text{QBF}_{2,\exists}$ 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 $\text{QBF}_{2,\exists}$, 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. $\square$

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 $\text{QBF}_{2,\exists}$ 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 propositions are entertained or believed (which amounts to being interested in whether arguments can be built for given propositions). For thoughtful and skeptical agents we need to consider whether an argument is undercut so that we can determine acceptability.

Proposition 25 *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) $v' \models H' \to h'$ and (ii) $v' \models \neg h \leftrightarrow h'$.

For hardness, there is a straightforward reduction from the $\text{QBF}_{2,\exists}$ 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 and determining acceptability (which will include showing that at least one argument has no undercutter) is computationally harder than establishing whether propositions are entertained or believed.

From these results we can see that, as it stands, a direct implementation of argumentation-based dialogues will be computationally very expensive. Indeed it will be intractable. Of course, as mentioned above, this is not surprising. Of more interest is the fact that we can home in on three separate areas which give rise to this intractability. First there is the construction of arguments; then there is the problem of determining if an argument, once constructed, is minimal; finally there is the problem of determining if there are undercutters for a given argument. Considering the proofs of the relevant results, it is clear that the key element as far as generating complexity is concerned is the dependence on establishing proof—it is the co-NP-completeness of establishing proof in propositional logic that raises the complexity so high in the polynomial hierarchy. What we need to do next is to look at using languages which have more efficient mechanisms for establishing proof with our approach (it is possible to establish proof in propositional Horn clauses, for example, in polynomial time) and see how that affects the complexity of argumentation-based dialogue. Of course, more efficient languages are typically less expressive, and future work will concentrate on establishing the tradeoffs in a similar manner to that in [46].

8 Related work

In the last few years, the formal study of argumentation became a hot topic in Artificial Intelligence, in particular in the area of multi-agent systems and in nonmonotonic and uncertain reasoning. In nonmonotonic and uncertain reasoning, argumentation systems have been used to define inference systems for existing nonmonotonic logics, as in the work of Geffner [12], or to define non-standard (most often nonmonotonic) consequence relations for a particular logic based on some notion of argument. In this latter line of work, that of Dung [11] has been particularly influential (not least upon the development of the approach we base our work on [1]), and has echoes in the work of Prakken and Sartor [35] and Vreeswijk [44]. Many more approaches are surveyed in [36] and [7].

On the multi-agent systems side, there has also been a good deal of work on argumentation and we will only discuss the most relevant examples. While the first work that we are aware of in the mainstream agent literature is that of Sycara [42, 43], many of the concepts and problems were studied simultaneously in the field of Artificial Intelligence and Law. Particularly important in showing the scope of systems of argumentation was Gordon's work on The Pleadings Game [13] and Zeno [14] and subsequent refinements of these ideas proposed by Loui [21] and Prakken [33, 34]. There has also been much relevant work in the area of natural language processing, though, as argued before, this deals with a much more complex task than one in which languages and agents can be engineered to carry out simple dialogues, and therefore a good deal of the work in natural language dialogues is not directly relevant to our work. However, the work of Grosz [15, 16] deserves a mention to indicate the ancestry of work in natural language on argumentation, and the fact that it has covered topics such as shared plans which our work has not. Reed's work [37] is also important for having brought the work of Walton and Krabbe to the attention of the agents community, and also for having provided a framework for combining different types of dialogue (later extended by [25]).

We distinguish the work in this paper from that of others in the literature first of all by its scope. Much of the existing work has dealt with a particular kind of dialogue, a form of deliberation in the case of Dignum and colleagues [8, 9] (though deliberation with strong overtones of persuasion), persuasion in the work of Prakken [33, 34], Gordon [13, 14] and Loui [21], and negotiation (though again with strong overtones of persuasion and also deliberation) in the work of Schroeder [40] and some of our previous work [28, 30]. In spirit our work is close to that of Sadri *et al.* [39] who have looked at the termination of negotiation dialogues (again having overtones of deliberation in the terminology of Walton and Krabbe) and the effect of agent types [38] and Kakas and Toni [20] who have looked at the complexity of some types of argumentation. We also see it as a continuation of our previous work on the complexity of negotiation dialogues [46].

However, the main way in which our work differs from that which has come previously is that, so far as we are aware, we are the first to have tried to identify a common framework in which a number of different kinds of dialogue can be expressed, the first to specify protocols for a range of dialogue types, and the first to study the properties of these protocols in such detail. The protocols we have studied here are very simple, but this was a conscious choice which was made in order to simplify the tasks at hand (not least as a result of trying to obtain similar results for the marginally more complex dialogues of [5]), and one that we believe to be justified by the fact that we have, for instance, exposed some interesting behaviour in these dialogues.

9 Conclusions

This paper has examined three types of argumentation-based dialogue between agents—information seeking, inquiry and persuasion, from the typology of [45]—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 in a reasonable number of steps, 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 [3] by precisely defining a set of protocols (albeit quite rigid ones) and a range of agent attitudes (in [3] 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—some initial results on this are presented in [4]—examining the effect of adding new locutions (such as *retract*) to the language, and identifying additional properties of the dialogues (such as the the extent to which the protocols defined here place restrictions on the outcomes of dialogues given what agents have in their knowledge-bases). We are currently investigating these matters along with further dialogue types—negotiation and deliberation in the typology of [45] and planning dialogues [15]— as well as more complex kinds of the dialogue types studied here, and additional complexity issues (including the use of languages other than propositional logic). Preliminary thoughts on the outcome-related properties may be found in [32], and an analysis of a simple form of deliberation dialogue is given in [29].

Another point that we are beginning to work on is the question of how the knowledge-base of an agent evolves through both an individual dialogue (are there ways that an agent should increase its knowledge in addition to accepting propositions asserted by another agent?) and across a number of dialogues (what kind of belief revision is appropriate at the end of a dialogue?).

Acknowledgments

This work was partly funded by the EU funded Project IST-1999-10948, and the NSF funded project REC-02-19347. We are extremely grateful for the detailed comments provided by the anonymous reviewers. These helped us to greatly improve the paper.

References

[1] L. Amgoud. *Contribution a l'integration des préferences dans le raisonnement argumentatif.* PhD thesis, Université Paul Sabatier, Toulouse, July 1999.

[2] 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.

[3] 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.

[4] L. Amgoud, N. Maudet, and S. Parsons. An argumentation-based semantics for agent communication languages. In *Proceedings of the 15th European Conference on Artificial Intelligence*, 2002.

[5] 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.

[6] 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.

[7] D. V. Carbogim, D. Robertson, and J. Lee. Argument-based applications to knowledge engineering. *The Knowledge Engineering Review*, 15:119–149, 2000.

[8] 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.

[9] F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Creating collective intention through dialogue. In J. Cunningham and D. Gabbay, editors, *Proceedings of the International Conference on Formal and Applied Practical Reasoning*, pages 145–158, London, UK, 2000. Department of Computing, Imperial College, University of London.

[10] J. Doyle and R. H. Thomason. Background to qualitative decision theory. *AI Magazine*, 20(2):55–68, 1999.

[11] 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.

[12] H. Geffner and J. Pearl. Conditional entailment: Bridging two approaches to default reasoning. *Artificial Intelligence*, 53:209–244, 1992.

[13] T. F. Gordon. The Pleadings Game: An exercise in computational dialectics. *Artificial Intelligence and Law*, 2:239–292, 1994.

[14] T. F. Gordon and N. Karacapilidis. The Zeno argumentation framework. In *Proceedings of the 6th International Conference on AI and Law*, pages 10–18. ACM Press, 1997.

[15] 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.

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

[17] C. L. Hamblin. *Fallacies*. Methuen and Co Ltd, London, UK, 1970.

[18] C. Hitchens. *The Trial of Henry Kissinger*. Verso, London, UK, 2001.

[19] 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.

[20] K. C. Kakas and F. Toni. Computing argumentation in logic programming. *Journal of Logic and Computation*, 9:113–131, 1999.

[21] R. P. Loui. Process and policy: Resource-bounded non-demonstrative reasoning. *Computational Intelligence*, 14:1–38, 1998.

[22] J. D. MacKenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133, 1979.

[23] 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.

[24] 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.

[25] P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *Journal of Logic, Language, and Information*, 11(3):315–334, 2002.

[26] C. H. Papadimitriou. *Computational Complexity*. Addison Wesley, 1994.

[27] 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.

[28] 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.

[29] S. Parsons and P. McBurney. Argumentation-based dialogues for agent co-ordination. *Group Decision and Negotiation*, 12(3), 2003.

[30] 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.

[31] S. Parsons, M. Wooldridge, and L. Amgoud. An analysis of formal interagent dialogues. In C. Castelfranchi and W. L. Johnson, editors, *Proceedings of the First International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 394–401, New York, USA, 2002. ACM Press.

[32] S. Parsons, M. Wooldridge, and L. Amgoud. On the outcomes of formal inter-agent dialogues. In T. Sandholm and M. Yakoo, editors, *Proceedings of the Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, New York, USA, 2003. ACM Press.

[33] H. Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In *Proceedings of the Seventh European Workshop on Logic in Artificial Intelligence*, Berlin, Germany, 2000. Springer Verlag.

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

[35] H. Prakken and G. Sartor. Modelling reasoning with precedents in a formal dialogue game. *Artificial Intelligence and Law*, 6:231–287, 1998.

[36] H. Prakken and G. Vreeswijk. Logics for defeasible argumentation. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 218–319. Kluwer, Dordrecht, 2002.

[37] 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.

[38] F. Sadri, F. Toni, and P. Torroni. Dialogues for negotiation: Agent varieties and dialogue sequences. In *Proceedings of the International Workshop on Agent Theories, Architectures and Languages, ATAL*, 2001.

[39] F. Sadri, F. Toni, and P. Torroni. Logic agents, dialogues and negotiation: an abductive approach. In M. Schroeder and K. Stathis, editors, *Proceedings of the Symposium on Information Agents for E-Commerce*, York, UK, 2001. AISB.

[40] 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.

[41] M. P. Singh. A social semantics for agent communication languages. In *Proceedings of the IJCAI'99 Workshop on Agent communication languages*, pages 75–88, 1999.

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

[43] K. Sycara. Persuasive argumentation in negotiation. *Theory and Decision*, 28:203–242, 1990.

[44] G. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.

[45] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany, NY, USA, 1995.

[46] 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.