The mechanics of some formal inter-agent dialogues

Simon Parsons[1,2], Peter McBurney[2], Michael Wooldridge[2]

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

[2] Department of Computer Science, University of Liverpool,
Chadwick Building, Liverpool L69 7ZF, UK.
{p.j.mcburney,m.j.wooldridge}@csc.liv.ac.uk

Abstract. This paper studies argumentation-based dialogues between agents. It takes a previously defined system by which agents can trade arguments and examines in detail what locutions are passed between agents. This makes it possible to identify finer-grained protocols than has been previously possible, exposing the relationships between different kinds of dialogue, and giving a deeper understanding of how such dialogues could be automated.

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: to resolve differences of opinion and conflicts of interest; to work together to resolve dilemmas or find proofs; or simply to inform each other of pertinent facts. Many of these communication requirements cannot be fulfilled by the exchange of single messages. Instead, the agents concerned need to be able to exchange a sequence of messages which all bear upon the same subject. In other words they need the ability to engage in dialogues. As a result of this requirement, there has been much work on providing agents with the ability to hold such dialogues. Recently some of this work has considered argument-based approaches to dialogue, for example the work by Dignum *et al.* [5], Parsons and Jennings [17], Reed [24], Schroeder *et al.* [25] and Sycara [26].

Reed's work built on an influential model of human dialogues due to argumentation theorists Doug Walton and Erik Krabbe [27], and we also take their dialogue typology as our starting point. Walton and Krabbe set out to analyse the concept of commitment in dialogue, so as to "provide conceptual tools for the theory of argumentation" [27, page ix]. This led to a focus on persuasion dialogues, and their work presents formal models for such dialogues. In attempting this task, Walton and Krabbe recognised the need for a characterisation of dialogues, and so they present a broad typology for inter-personal dialogue. They make no claims for its comprehensiveness.

Their categorisation identifies six primary types of dialogues and three mixed types. The categorisation is based upon: what information the participants each have at the commencement of the dialogue (with regard to the topic of discussion); what goals the individual participants have; and what goals are shared by the participants, goals we may view as those of the dialogue itself. This *dialogue game* view of dialogues, revived by Hamblin [12] and extending back to Aristotle, overlaps with work on conversational policies (see, for example, [4, 7]), but differs in considering the entire dialogue rather than dialogue segments.

As defined by Walton and Krabbe, the three types of dialogue we consider here are:

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

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

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

Our previous work investigated capturing these types of dialogue using a formal model of argumentation [2], the protocols behind these types of dialogue, and properties and complexity the dialogues [20, 22], and the range of possible outcomes from the dialogues [21]. Here we extend this investigation, turning to consider the internal detail of the dialogues, detail that we have previously skated over.

There are two reasons why we do this. First, we want to make sure that the protocols we introduced in [20] are fully specified. From our previous work, we already know that they capture the essence of information seeking, inquiry and persuasion—here we aim to ensure that all the necessary mechanics are in place as well. Second, our previous analysis suggests some deep connections between the different protocols—they seem to be variations on a theme rather than separate themes—and looking at their internal detail is one way to find out if these connections exist.

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 Grosz and Sidner [11] for example, we choose to focus in this way in order to simplify our task—dealing with artificial languages avoids much of the complexity inherent in natural language dialogues.

2 Background

In this section we briefly introduce the formal system of argumentation, due to Amgoud [1], that forms the backbone of our approach. This is inspired by the work of Dung [6] but goes further in dealing with preferences between arguments. Further details are available in [1]. We start with a possibly inconsistent knowledge base Σ with no deductive closure. We assume Σ contains formulas of a propositional language $\mathcal{L}$. $\vdash$ stands for classical inference, $\rightarrow$ for material implication, 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 proper subset of H satisfying both 1. and 2. exists.

H is called the *support* of A, written $H = \text{Support}(A)$ and h is the *conclusion* of A written $h = \text{Conclusion}(A)$.

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 $\exists h \in \mathit{Support}(A_2)$ such that $h \equiv \neg \mathit{Conclusion}(A_1)$.

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

To capture the fact that some facts are more strongly believed[1] we assume that any set of facts has a preference order over it. We suppose that this ordering derives from the fact that the knowledge base Σ is stratified into non-overlapping sets $\Sigma_1, \ldots, \Sigma_n$ such that facts in Σ_i are all equally preferred and are more preferred than those in Σ_j where $j > i$. The preference level of a nonempty subset H of Σ, $\mathit{level}(H)$, 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*, $\mathit{Pref}(A_1, A_2)$, iff $\mathit{level}(\mathit{Support}(A_1)) \leq \mathit{level}(\mathit{Support}(A_2))$.

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

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

[1] Here we only deal with beliefs, though the approach can also handle desires and intentions as in [19] and could be extended to cope with other mental attitudes.

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

- $\mathcal{A}(\Sigma)$ 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.*
- A set of arguments $\mathcal{S}$ *defends* A iff: $\forall$ B undercuts A and A does not defend itself against B then $\exists$ $C \in \mathcal{S}$ such that C undercuts B and B does not defend itself against C.

Henceforth, $C_{Undercut,Pref}$ will gather all non-undercut arguments and arguments defending themselves against all their undercutting arguments. In [1], Amgoud showed that the set $\underline{\mathcal{S}}$ 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:

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

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

An acceptable argument is one which is, in some sense, proven since all the arguments which might undermine it are themselves undermined. However, this status can be revoked following the discovery of a new argument (possibly as the result of the communication of some new information from another agent).

3 Locutions and attitudes

As in our previous work, agents decide what they know by determining which propositions they have acceptable arguments for. They assert propositions for which they have acceptable arguments, and accept propositions put forward by other agents if they find that the arguments are acceptable to them. 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, for example called P and C. Each agent has a knowledge base, Σ_P and Σ_C respectively, containing their beliefs. In addition, each agent has a further knowledge base, accessible to both agents, containing commitments made in the dialogue[3]. These commitment stores are denoted $CS(P)$ and $CS(C)$ respectively, and in this dialogue system everyting in an agent's commitment store is either in its knowledge base, or in the other agent's commitment store. Note that the union of the commitment stores can be viewed as the state of the dialogue at a given time. Since each agent has access to its own private knowledge base and both commitment stores, P can make use of $\langle \mathcal{A}(\Sigma_P \cup CS(C)), \mathit{Undercut}, \mathit{Pref} \rangle$[4] and C can make use of $\langle \mathcal{A}(\Sigma_C \cup CS(P)), \mathit{Undercut}, \mathit{Pref} \rangle$.

All the knowledge bases contain propositional formulas, are not closed under deduction, and are stratified by degree of belief as discussed above. Here we assume that these degrees of belief are static and that both the players agree on them, though it is possible [3] to combine different sets of preferences, and it is also possible to have agents modify their beliefs on the basis of the reliability of their acquaintances [16].

With this background, we can present the set of dialogue moves first introduced in [20]. Each locution has a rule describing how to update commitment stores after the move, and groups of moves have conditions under which the move can be made—these are given in terms of the agents' assertion and acceptance attitudes (defined below). For all moves, player P addresses the ith move of the dialogue to player C.

assert(p) where p is a propositional formula.

$$CS_i(P) = CS_{i-1}(P) \cup \{p\} \text{ and } CS_i(C) = CS_{i-1}(C)$$

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

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

$$CS_i(P) = CS(P)_{i-1} \cup S \text{ and } CS_i(C) = CS_{i-1}(C)$$

The counterpart of these moves are the acceptance moves. They can be used whenever the protocol and the agent's acceptance attitude allow.

accept(p) p is a propositional formula.

$$CS_i(P) = CS_{i-1}(P) \cup \{p\} \text{ and } CS_i(C) = CS_{i-1}(C)$$

[3] Following Hamblin [12] commitments here are propositions that an agent is prepared to defend.

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

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

$CS_i(P) = CS_{i-1}(P) \cup S$ 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.

$CS_i(P) = CS_{i-1}(P)$ and $CS_i(C) = CS_{i-1}(C)$

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

question(p) where p is a propositional formula.

$CS_i(P) = CS_{i-1}(P)$ and $CS_i(C) = CS_{i-1}(C)$

Finally, we have the locution *reject*(p), which does not change the commitment stores, but indicates that the agent which utters it will not *accept* p.

We refer to this set of moves as the set $\mathcal{M}'_{DC}$. The locutions in $\mathcal{M}'_{DC}$ are similar to those discussed in models of legal reasoning [8, 23] and it should be noted that there is no *retract* locution. Note that these locutions are ones used within dialogues—locutions such as those discussed in [15] would be required to frame dialogues.

We also need to define the attitudes which control the assertion and acceptance of propositions.

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

- a *confident* agent can assert any proposition p for which it can construct an argument (S, p).
- a *careful* agent can assert any proposition p for which it can construct an argument, if it is unable to construct a stronger argument for $\neg p$.
- a *thoughtful* agent can assert any proposition p for which it can construct an acceptable argument (S, p).

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

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

Since agents are typically involved in both asserting and accepting propositions, we denote the combination of an agent's two attitudes as

$$\langle \textit{assertion attitude} \rangle / \langle \textit{acceptance attitude} \rangle$$

The effects of this range of agent attitudes on dialogue outcomes is studied in [22], and for the rest of this paper we will largely ignore agents' attitudes, though the distinction between agents that are credulous and those that are not becomes important in a couple of places.

4 Types of dialogue

Previously [20] we defined three protocols for information seeking, inquiry and persuasion dialogues. These protocols are deliberately simple, the simplest we can imagine that can satisfy the definitions given by [27], since we believe that we need to understand the behaviour of these simple protocols before we are to able to understand more complex protocols.

4.1 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, the other agent is B, and the proposition that the dialogue is concerned with is p, then the dialogue starts with A having no argument for p or $\neg p$, and one possible protocol for conducting an information-seeking dialogue about p is the following protocol we denote as $\mathcal{IS}$:

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

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

4.2 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 [14]). 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 [15], and from this point can adopt the following protocol $\mathcal{I}$:

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

[5] The protocol in [2] allows an agent to interject with $question(p)$ for any p at several points, making it possible for a dialogue between two agents to continue indefinitely.

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

Here the initial conditions of the dialogue are that neither agent has an argument for p.

This protocol[6] is basically a series of implied $\mathcal{IS}$ dialogues. First A asks "do you know of anything which would imply p were it known?". B replies with one, or the dialogue terminates with $\mathcal{U}$. If A accepts the implication, B asks "now, do you know q, or any r which would imply q were it known?", and the process repeats until either the process bottoms out in a proposition which both agents agree on, or there is no new implication to add to the chain.

4.3 Persuasion

In a persuasion dialogue, one party seeks to persuade another agent to adopt a belief or point-of-view it does not currently hold. In other words, the dialogue starts with one agent having an argument for a proposition p, and the other either having no argument for p, or having an argument for $\neg p$[7]. The dialogue game DC, on which the moves in [2] are based, is fundamentally a persuasion game, so the protocol below results in games which are very like those described in [2]. This protocol, $\mathcal{P}$, is as follows, where agent A is trying to persuade agent B to accept p.

1. A *asserts* p.
2. B *accepts* p if its acceptance attitude allows, if not B *asserts* $\neg p$ if it is allowed to, or otherwise *challenges* 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 *challenged*, then:
   (a) A asserts S, the support for p;
   (b) Goto 2 for each $s \in S$ in turn.

[6] Which differs from the inquiry dialogue in [20] in the *accept* moves in step 6.

[7] This second condition is better stated as having an argument for $\neg p$ that is acceptable according to its acceptability attitude, and no argument for p that is acceptable in this way.

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

We should point out that this kind of persuasion dialogue does not assume that agents necessarily start from opposite positions, one believing p and one believing $\neg p$. Instead one agent believes p and the other may believe $\neg p$, but also may believe neither p nor $\neg p$. This is perfectly consistent with the notion of persuasion suggested by Walton and Krabbe [27].

Note that all three of these protocols have the same core steps. One agent *assert*s something, the other *accept*s if it can, otherwise it *challenge*s. A *challenge* provokes the *assert*ion of the grounds, which are in turn either *accept*ed or *challenge*d. The proposition p that is the first assertion, and the central proposition of the dialogue, is said to be the *subject* of the dialogue. This basic framework has been shown [18, 20] to be capable of capturing a range of dialogue types, and we have studied a number of the properties of these dialogues including termination and complexity [22] and what their possible outcomes are [21]. Our purpose here is to look in more detail at the structure of these dialogues, in particular the core steps.

5 Dialogue mechanics

As already mentioned, the dialogue protocols given above have the same core steps, and it is interesting to consider these steps as forming *atomic protocols* from which other protocols are constructed. Are these truly atomic, in the sense that they cannot be broken down into combinations of simpler protocols? What combinations of atomic protocols make sense (in other words, are there protocols that we have not yet identified which can be made from the atomic protocols)? Indeed, we haven't as yet even answered the most basic question—what atomic protocols are there?

5.1 Identifying atomic protocols

To identify what atomic protocols there are, we will start by writing out a complete $\mathcal{IS}$ dialogue. We imagine a dialogue between agent A and agent B about a proposition p. A typical $\mathcal{IS}$ dialogue might proceed as follows, where the dialogue has an acceptance outcome of p for B (so that B *assert*s p, and A later accepts it[8]). This dialogue is in what we will call *extensive form*, by which we denote the fact that every choice of locution is such that it tends to extend the dialogue as much as possible, so what we have here is the longest possible dialogue that can arise. Clearly any of A's *accept*s could equally well be a *reject*, and the dialogue would stop after at most two more *reject*s (and, indeed, after one *reject*, *reject* would be the only locution that could possibly be uttered).

[8] The notion of an acceptance outcome is formally defined in [21].

```
A: $question(p)$
B: $assert(p)$
    A: $challenge(p)$
    B: $assert\ (\bigcup_i \{s_i\}_{i=1\ldots n})$
        A: $challenge(s_1)$
        B: $assert(\{s_1\})$
        A: $accept(s_1)$
        A: $challenge(s_2)$
        B: $assert(\{s_2\})$
        A: $accept(s_2)$

        ⋮

        A: $challenge(s_n)$
        B: $assert(\{s_n\})$
        A: $accept(s_n)$
    A: $accept\ (\bigcup_i \{s_i\}_{i=1\ldots n})$
A: $accept(p)$
```

If we consider this dialogue to be made up of a series of (indented) *sub-dialogues*—each of which is an instantiation of an atomic protocol—we can easily identify two distinct atomic protocols. We first have Q (for "question") protocol:

A: $question(x)$
B: $assert(y)$
A: $accept(y)$ or $reject(y)$

where y is either x or $\neg x$. In the example above, the outermost sub-dialogue is built according to this protocol, and every other sub-dialogue is embedded in this sub-dialogue.

This Q protocol is too simple to even describes simplest possible kind of $\mathcal{IS}$ dialogue on its own, since the only kind of dialogue it covers is one in which A asks the question, B replies, and A immediately accepts.

Proposition 9. *A dialogue under $\mathcal{IS}$ with subject p between agents A and B will never involve only a dialogue under the Q atomic protocol.*

Looking again at the example dialogue above, we can identify instantiations of a second atomic protocol, which we will call A. This is the protocol responsible for the "core steps" mentioned at the end of the last section (though without the initial *assert* that at first seems to be an obvious part of those steps). In other words A is:

A: $challenge(x)$
B: $assert(y)$
A: $accept(y)$ or $reject(y)$

In the outermost instantiation of this protocol in the example, x is the last proposition to be *assert*ed and y is the set of propositions that form the support of x. The dialogue generated by this instantiation of the atomic protocol is embedded within the dialogue generated by the Q protocol[9], and then has n A dialogues nested within it. All the other instantiations of the A protocol are nested within the first A dialogue, follow each other in sequence, and all have the same form. x is one of the propositions s_i in the set *assert*ed in the first A dialogue, and y is the set $\{s_i\}$ which is the only possible set of grounds supporting the assertion s_i.

Before proceeding any further, it is clear from this exposition of a dialogue under $\mathcal{IS}$ that any such dialogue will terminate provided that the set of grounds S_i is finite, and that it will terminate in time proportional to $|S_i|$. In other words:

Proposition 10. *A dialogue under $\mathcal{IS}$ with subject p between agents A and B, in which will A utters the first illocution, will terminate in at most $O(|S_i|)$ steps, where B has an argument (S_i, p) or $(S_i, \neg p)$.*

This is an even tighter bound on the length of the dialogue than we obtained in [22], and doesn't depend upon the knowledge base Σ_B of the agent making the initial assertion being finite.

Having identified the atomic protocols underlying $\mathcal{IS}$, we can turn to look at $\mathcal{I}$. As described above, a full-fledged $\mathcal{I}$ dialogue would look as follows:

$$\begin{array}{l}
\text{A: } \mathit{assert}(q \rightarrow p) \\
\quad \text{B: } \mathit{challenge}(q \rightarrow p) \\
\quad \text{A: } \mathit{assert}\left(\bigcup_i \{s_i\}_{i=1\ldots n}\right) \\
\quad\quad \text{B: } \mathit{challenge}(s_1) \\
\quad\quad \text{A: } \mathit{assert}(\{s_1\}) \\
\quad\quad \text{B: } \mathit{accept}(s_1) \\
\quad\quad\quad \vdots \\
\quad\quad \text{B: } \mathit{challenge}(s_n) \\
\quad\quad \text{A: } \mathit{assert}(\{s_n\}) \\
\quad\quad \text{B: } \mathit{accept}(s_n) \\
\quad \text{B: } \mathit{accept}\left(\bigcup_i \{s_i\}_{i=1\ldots n}\right) \\
\text{B: } \mathit{accept}(q \rightarrow p) \\
\text{B: } \mathit{assert}(r \rightarrow q) \\
\quad \vdots \\
\text{A: } \mathit{accept}(r \rightarrow q) \\
\quad \vdots \\
\text{A: } \mathit{assert}(v \rightarrow w) \\
\quad \vdots \\
\text{B: } \mathit{accept}(v \rightarrow w) \\
\text{B: } \mathit{assert}(v) \\
\quad \vdots
\end{array}$$

[9] From here on, we will refer to "the dialogue generated by the X sub-protocol" as the "X dialogue", where this usage is not ambiguous and in Section 5.2 we will develop this idea formally.

A: $\mathit{accept}(v)$
A: $\mathit{accept}(p)$
B: $\mathit{accept}(p)$

In other words, $\mathcal{I}$ start with an agent asserting a formula $q \rightarrow p$ that provides a means to infer p, the subject of the dialogue. The agents then engage in the same kind of nested A dialogue we saw above to determine if this formula is an acceptance outcome for the first agent. Then the second agent asserts a formula that provides the means to infer q (and so is another step in the proof of p). This process continues until one agent accepts the latest step in this chain, $v \rightarrow w$, and then can get the other agent to accept v.

This analysis exposes a number of weaknesses with the $\mathcal{I}$ protocol, which we have already noted [22][10]. One such weakness is the rigidity of the protocol—it relies on strict turn taking by the agents, they have to supply sequential pieces of the proof, and it explores only one possible proof of the subject[11]. Another weakness is the fact that it assumes the agents have already agreed to engage in an inquiry dialogue—unlike the information seeking dialogue, there is no initial illocution to specify "let's start trying to prove p".

Without such an utterance, the structure of an $\mathcal{I}$ dialogue isn't a combination of clearly identifiable atomic dialogues. It is perhaps more elegant to consider adding an utterance $\mathit{prove}(p)$ which has exactly the sense of "let's start trying to prove p" and imagine a P (for "proof") atomic dialogue which runs as either of

B: $\mathit{prove}(x)$
A: $\mathit{assert}(x)$
B: $\mathit{accept}(x)$ or $\mathit{reject}(x)$

or

B: $\mathit{prove}(x)$
A: $\mathit{assert}(y \rightarrow x)$
B: $\mathit{accept}(y \rightarrow x)$ or $\mathit{reject}(y \rightarrow x)$

Such a dialogue could produce a version of the $\mathcal{I}$ example above when iterated as:

B: $\mathit{prove}(p)$
A: $\mathit{assert}(q \rightarrow p)$
nested A dialogues about $q \rightarrow p$
B: $\mathit{accept}(q \rightarrow p)$
A: $\mathit{prove}(q)$
$\vdots$
B: $\mathit{assert}(v)$
nested A dialogues about v
A: $\mathit{accept}(v)$
B: $\mathit{accept}(p)$
A: $\mathit{accept}(p)$

[10] Another weakness that we have not mentioned before is that as it stands the $\mathcal{I}$ protocol only allows the construction of proofs that are chains of material implications. A more general formulation would require each assertion to be any formula which would help in the proof of p.

[11] [22] also provides some solutions to these particular problems.

Although such a dialogue is an extension of $\mathcal{I}$ as we have previously defined it (and requires an extension of the set of locutions to $\mathcal{M}'_{DC} \cup \{prove(p)\}$), this is what we will consider to be a prototypical inquiry dialogue, $\mathcal{I}''$, for the remainder of this paper[12]. Defining P in this way gives us analogous results to those for Q, for instance:

Proposition 11. *A dialogue under $\mathcal{I}''$ with subject p between agents A and B will never only involve a dialogue under the P atomic protocol.*

Finally, we can look for the atomic protocols that make up the $\mathcal{P}$ protocol for persuasion dialogues. As defined above, there are two ways that a persuasion dialogue may, in general, be played out. The simplest is as follows:

A: $assert(p)$
  B: $challenge(p)$
  A: $assert\left(\bigcup_i \{s_i\}_{i=1\ldots n}\right)$
    B: $challenge(s_1)$
    A: $assert(\{s_1\})$
    B: $accept(s_1)$
    B: $challenge(s_2)$
    A: $assert(\{s_2\})$
    B: $accept(s_2)$
    $\vdots$
    B: $challenge(s_n)$
    A: $assert(\{s_n\})$
    B: $accept(s_n)$
  B: $accept\left(\bigcup_i \{s_i\}_{i=1\ldots n}\right)$
B: $accept(p)$

This kind of dialogue, which we might call *persuasion*$_1$, is the kind which arises for example when B does not initially have an opinion about whether or not p is true. As a result, $\mathcal{P}$ generates a dialogue that has the same form as a $\mathcal{IS}$ dialogue though without the initial *question*. Just as in the case for the inquiry dialogue without the *prove* locution, the fact that this is a persuasion dialogue is implicit—any assertion can be the start of a $\mathcal{P}$ dialogue. To make the start of the dialogue explicit, we could insist that before A makes its initial *assert*, it signals the start of a persuasion dialogue by using a locution $know(p)$, which has the intended meaning "do you know that p is the case?". (Again we will have to extend the set of locutions, this time to $\mathcal{M}'_{DC} \cup \{prove(p), know(p)\}$, a set we will call $\mathcal{M}^{PK}_{DC}$.)

The other way that $\mathcal{P}$ (well, in fact it is a new protocol $\mathcal{P}'$ which includes the *know* locution) can play out is when B replies to the initial assertion of p with its own assertion of $\neg p$ in which case we get a *persuasion*$_2$ dialogue that looks like:

[12] And we should point out that modifying $\mathcal{I}$ in this way will not change any of the properties already proved for it.

A: $\mathit{know}(p)$
A: $\mathit{assert}(p)$
B: $\mathit{know}(\neg p)$
B: $\mathit{assert}(\neg p)$
A: $\mathit{challenge}(\neg p)$
B: $\mathit{assert}\left(\bigcup_i \{s_i\}_{i=1\ldots n}\right)$
A: $\mathit{challenge}(s_1)$
B: $\mathit{assert}(\{s_1\})$
A: $\mathit{accept}(s_1)$
A: $\mathit{challenge}(s_2)$
B: $\mathit{assert}(\{s_2\})$
A: $\mathit{accept}(s_2)$
$\vdots$
A: $\mathit{challenge}(s_n)$
B: $\mathit{assert}(\{s_n\})$
A: $\mathit{accept}(s_n)$
A: $\mathit{accept}\left(\bigcup_i \{s_i\}_{i=1\ldots n}\right)$
A: $\mathit{accept}(\neg p)$
B: $\mathit{reject}(p)$

From this we can identify a new atomic protocol:

A: $\mathit{know}(x)$
A: $\mathit{assert}(x)$
B: $\mathit{reject}(x)$ or $\mathit{accept}(x)$

which we will call K after its first locution. It is then clear that *persuasion*$_1$ is just the usual set of nested A dialogues within a K dialogue, and that *persuasion*$_2$ is a *persuasion*$_1$ nested within a further K. Exactly as for the A and Q protocols, the K protocol cannot generate a $\mathcal{P}'$ dialogue on its own:

Proposition 12. *A dialogue under $\mathcal{P}'$ with subject p between agents A and B will never involve only a dialogue under the* K *atomic protocol.*

5.2 Combinations of atomic protocols

We can formally describe how dialogues under $\mathcal{IS}$ $\mathcal{I}''$, and $\mathcal{P}'$ are constructed from the atomic protocols using the notation we developed in [15]. In that paper we defined (albeit for dialogues rather than protocols):

Iteration: If G is a protocol, then G^n is also a protocol, being that protocol which consists of the n-fold repetition of G, each occurrence being undertaken until closure, and then being followed immediately by the next occurrence.

Sequencing: If G and H are both protocols, then $G;H$ is also a protocol, representing that protocol which consists of undertaking G until its closure and then immediately undertaking H.

Parallelization: If G and H are both protocols, then $G \cap H$ is also a protocol, representing that protocol which consists of undertaking both G and H simultaneously, until each are closed.

Embedding: If G and H are both protocols, and Φ is a finite set of legal locution sequences in G, then $G[H|\Phi]$ is also a protocol, representing that protocol which consists of undertaking G until a sequence in Φ has been executed, and then switching immediately to protocol H which is undertaken until its closure, whereupon protocol G resumes from immediately after the point where it was interrupted and continues until closure. Protocol H is said to be embedded in G, at one level lower than G. In the time between when H opens and closes, protocol G remains open, no matter how many embedded protocols H itself may contain.

Up to this point it has sufficed to talk informally about dialogues generated by protocols, but for the remainder of the paper we need to be a bit more formal. We start by defining what a dialogue is:

Definition 13. A *dialogue* is an ordered sequence of locutions that are valid according to some protocol.

Clearly, a given protocol can clearly generate many different dialogues, with the exact dialogue being dependent upon which agents are involved (the important aspect being what the agents know), in what order the agents generate locutions (which is specified by which agent makes the first locution), and the subject of the dialogue. We can therefore fully specify a dialogue by identifying the protocol and these features. For instance, we write:

$$\mathsf{Q}^{B \rightarrow A}(\Sigma_B, \Sigma_A)(p)$$

to denote the dialogue generated by protocol Q, with subject p, between agents A and B, with knowledge bases Σ_A and Σ_B, where the first locution is uttered by B. If any of these specifiers have no bearing on a particular dialogue, we omit them.

With this notation, we can describe our first dialogue example as:

$$\mathsf{Q}^{A \rightarrow B}(p) \left[\mathsf{A}^{A \rightarrow B}(p) \left[\left(\mathsf{A}^{A \rightarrow B}(s_i) \right)^n | \{ assert(S) \} \right] | \{ assert(p) \} \right]$$

where (S, p) is an argument in $\mathcal{A}(\Sigma_B \cup CS(A))$, $S = \{s_1, \ldots, s_n\}$.

Now, while any information seeking dialogue won't necessarily be exactly the same as this, it will have exactly this form. To be able to express what "exactly this form" is, we need the following notion:

Definition 14. A protocol G *sequence includes* a protocol H if, for any two agents A and B, with knowledge bases Σ_A and Σ_B, G can generate all the dialogues that H can generate.

Thus $\mathcal{IS}$ sequence includes Q, but Q does not sequence include $\mathcal{IS}$ (because, for example, Q cannot generate dialogues like our first example on its own). Sequence inclusion gives us a notion of equivalence between protocols:

Definition 15. Two protocols G and H are *sequence equivalent* if G sequence includes H and H sequence includes G.

In other words two protocols are sequence equivalent if they generate exactly the same sets of dialogues. This is a new notion of equivalence between protocols, one that we didn't identify in [13], but it is close to the notion of bisimulation equivalence from that paper.

With these ideas, we can show a more precise version of Proposition 9:

Proposition 16. $\mathsf{Q}^{A \to B}(p)$ *does not sequence include* $\mathcal{IS}^{A \to B}(p)$.

In fact we can even drop the specifier p, since any dialogue that opens with a *question* will play out in the same way. This sets a lower limit on the complexity of a $\mathcal{IS}$ dialogue, in the sense that it must contain more than the locutions that can be generated by a single atomic protocol. In fact, it must contain at least two atomic protocols:

Proposition 17. *If A is credulous, then:*

$$\mathsf{Q}^{A \to B}(p) \left[\mathsf{A}^{A \to B}(p) | \{assert(p)\}\right]$$

where (S, p) is an argument in $\mathcal{A}(\Sigma_B \cup CS(A))$, will be sequence equivalent to $\mathcal{I}''^{A \to B}(p)$.

This dialogue is the simplest kind of information seeking dialogue that is possible under the $\mathcal{IS}$ protocol. If A isn't credulous, we need the full kind of dialogue in our first example to capture the $\mathcal{IS}$ protocol:

Proposition 18. *If A is not credulous, then:*

$$\mathsf{Q}^{A \to B}(p) \left[\mathsf{A}^{A \to B}(p) \left[\left(\mathsf{A}^{A \to B}(s_i)\right)^n | \{assert(S)\}\right] | \{assert(p)\}\right]$$

where (S, p) is an argument in $\mathcal{A}(\Sigma_B \cup CS(A))$, $S = \{s_1, \ldots, s_n\}$, will be sequence equivalent to $\mathcal{IS}^{A \to B}(p)$

We can obtain similar results for the other kinds of dialogue. For inquiry dialogues we have a similar lower limit on the complexity of a dialogue:

Proposition 19. $\mathsf{P}^{A \to B}(p)$ *does not sequence include* $\mathcal{I}''^{A \to B}(p)$.

If A is credulous, then we get the simplest kind of $\mathcal{I}''$ dialogue:

Proposition 20. *If A is credulous, then:*

$$\mathsf{P}^{A \to B}(p) \left[\mathsf{A}^{A \to B}(p) | \{assert(p)\}\right]$$

where (S, p) is an argument in $\mathcal{A}(\Sigma_B \cup CS(A))$, will be sequence equivalent to $\mathcal{I}''^{A \to B}(p)$.

If A is not credulous, then the dialogue gets more complex. Exactly how complex is determined by the length of the proof assembled by the two agents.

Proposition 21. *If A is not credulous, then:*

$$\begin{array}{l}
\mathsf{P}^{A\rightarrow B}(p)\left[\mathsf{A}^{A\rightarrow B}(p)\left[\left(\mathsf{A}^{A\rightarrow B}(s_{i_j})\right)^n|\{assert(S_1)\}\right]|\{assert(p)\}\right]; \\
\mathsf{P}^{B\rightarrow A}(v_1)\left[\mathsf{A}^{B\rightarrow A}(v_1)\left[\left(\mathsf{A}^{B\rightarrow A}(s_{i_j})\right)^n|\{assert(S_2)\}\right]|\{assert(v_1)\}\right]; \\
\vdots \\
\mathsf{P}^{A\rightarrow B}(v_n)\left[\mathsf{A}^{A\rightarrow B}(v_n)\left[\left(\mathsf{A}^{A\rightarrow B}(s_{i_j})\right)^n|\{assert(S_n)\}\right]|\{assert(v_n)\}\right];
\end{array}$$

where (S_i, v_i) is an argument in $\mathcal{A}(\Sigma_B \cup CS(A))$, $S_i = \{s_{i_1}, \ldots, s_{i_n}\}$ for odd i, and (S_k, v_k) is an argument in $\mathcal{A}(\Sigma_A \cup CS(B))$, $S_k = \{s_{k_1}, \ldots, s_{k_n}\}$ for even k, such that $\{v_1, \ldots, v_n\} \vdash p$, will be sequence equivalent to $\mathcal{IS}^{A\rightarrow B}(p)$

Proof. □

This makes the iterative structure of inquiry dialogues clear, as well as the similarity between a single iteration and an $\mathcal{IS}$ dialogue.

For persuasion dialogues we have to consider two cases, *persuasion*$_1$ and *persuasion*$_2$, but the first two results hold for both kinds:

Proposition 22. $\mathsf{K}^{A\rightarrow B}(p)$ *does not sequence include* $\mathcal{P}'^{A\rightarrow B}(p)$.

The second result depends on B (since it is the agent to whom the assertion is made) rather than A as in the previous dialogues.

Proposition 23. *If B is credulous, then:*

$$\mathsf{K}^{A\rightarrow B}(p)\left[\mathsf{A}^{B\rightarrow A}(p)|\{assert(p)\}\right]$$

where (S, p) is an argument in $\mathcal{A}(\Sigma_A \cup CS(B))$, will be sequence equivalent to $\mathcal{P}'^{A\rightarrow B}(p)$.

It seems odd that this should hold for a *persuasion*$_2$ dialogue, since we know that in a *persuasion*$_2$ dialogue which starts with A uttering a *know*, B has an argument for $\neg p$. However, that is the nature of credulous agents—they accept anything backed by an argument. When the persuadee is not credulous, then we have to consider *persuasion*$_1$ and *persuasion*$_2$ dialogues separately. For a *persuasion*$_1$ dialogue we get a result just like that for $\mathcal{IS}$:

Proposition 24. *If B is not credulous, and $(S', \neg p) \notin \mathcal{A}(\Sigma_B \cup CS(A))$ then:*

$$\mathsf{K}^{A\rightarrow B}(p)\left[\mathsf{A}^{B\rightarrow A}(p)\left[\left(\mathsf{A}^{B\rightarrow A}(s_i)\right)^n|\{assert(S)\}\right]|\{assert(p)\}\right]$$

where (S, p) is an argument in $\mathcal{A}(\Sigma_A \cup CS(B))$, $S = \{s_1, \ldots, s_n\}$, will be sequence equivalent to $\mathcal{P}'^{A\rightarrow B}(p)$

However, it is easy to see that there is an important difference between this kind of dialogue and an $\mathcal{IS}$ dialogue, other than the first atomic dialogue, which is that the order in which the agents utter locutions is different. For a *persuasion*$_2$ dialogue we have:

Proposition 25. *If B is not credulous, and $(S', \neg p) \in \mathcal{A}(\varSigma_B \cup CS(A))$, then*

$$\mathsf{K}^{A\rightarrow B}(p)\left[\mathsf{K}^{B\rightarrow A}(\neg p)\right.$$
$$\left.\left[\mathsf{A}^{A\rightarrow B}(p)\left[\left(\mathsf{A}^{A\rightarrow B}(s_i)\right)^n|\{assert(S')\}\right]|\{assert(\neg p)\}\right]|\{assert(p)\}\right]$$

where $S' = \{s_1, \ldots, s_n\}$, will be sequence equivalent to $\mathcal{P}'^{A\rightarrow B}(p)$

These results, then, guarantee that the atomic protocols exactly capture, in a strong sense, the protocols we first identified in [20]. There is more that we have done concerning atomic protocols, for example examining legal combinations of them other than those given above, which we do not have room to include here, and making use of the specifications of the types of dialogue given above (for example to make formal comparisons of them). We will report these results in a later paper.

6 Conclusions

This paper has extended the analysis of formal inter-agent dialogues that we began in [20, 21]. The main contribution of this paper is to identify a set of atomic protocols which can be combined (in the ways that we described in [15]) to give exactly the protocols introduced in [20, 21]. In this way we have done what we said we would in the introduction, putting the protocols from our previous work under the microscope to find out exactly how they work. As a result of this work, we now have a precise formal characterisation of our protocols, and are now in a position to start to compare protocols in some of the ways we suggested in [13].

One thing that has emerged, rather to our surprise, is a link between our work and conversation policies (for example [9]). Though we have yet to look at the matter in detail, it seems to us that the atomic protocols we have identified here are rather like conversation policies as we understand them—rules about short sequences of locutions which assemble sections of an overall conversation between agents. We intend to look at this matter more in the near future, however, it seems that we can think of the kinds of protocol we have been studying as composed of combinations of conversation policies, something that suggests it will be particularly important to establish the full range of sensible combinations of atomic protocols.

More work, of course, remains to be done in this area in addition to that outlined above. Particularly important are: determining the relationship between the locutions we use in these dialogues and those of agent communication languages such as the FIPA ACL; examining the effect of adding new locutions (such as *retract*) to the language; extending the system with a more detailed model of preferences; and providing an implementation. We are currently investigating these matters along with further dialogue types, such as planning dialogues [10].

Acknowledgments

This work was partially supported by NSF REC-02-19347. Many thanks to Eva Cogan for pointing out a number of typos in the published version of this paper, that are corrected in this version.

References

1. L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation framework. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence*, pages 1–7, 1998.
2. L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In E. Durfee, editor, *Proceedings of the Fourth International Conference on Multi-Agent Systems*, pages 31–38, Boston, MA, USA, 2000. IEEE Press.
3. L. Amgoud and S. Parsons. Agent dialogues with conflicting preferences. In J.-J. Meyer and M. Tambe, editors, *Proceedings of the 8th International Workshop on Agent Theories, Architectures and Languages*, pages 1–15, 2001.
4. B. Chaib-Draa and F. Dignum. Trends in agent communication language. *Computational Intelligence*, 18(2):89–101, 2002.
5. F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Agent theory for team formation by dialogue. In C. Castelfranchi and Y. Lespérance, editors, *Seventh Workshop on Agent Theories, Architectures, and Languages*, pages 141–156, Boston, USA, 2000.
6. P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence*, 77:321–357, 1995.
7. R. A. Flores and R. C. Kremer. To commit or not to commit. *Computational Intelligence*, 18(2):120–173, 2002.
8. T. F. Gordon. The pleadings game. *Artificial Intelligence and Law*, 2:239–292, 1993.
9. M. Greaves, H. Holmback, and J. Bradshaw. What is a conversation policy? In F. Dignum and M. Greaves, editors, *Issues in Agent Communication*, Lecture Notes in Artificial Intelligence 1916, pages 118–131. Springer, Berlin, Germany, 2000.
10. 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.
11. B. J. Grosz and C. L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.
12. C. L. Hamblin. *Fallacies.* Methuen and Co Ltd, London, UK, 1970.
13. M. W. Johnson, P. McBurney, and S. Parsons. When are two protocols the same? In M-P Huget, editor, *Communication in Multi-Agent Systems: Agent Communication Languages and Conversation Policies*, Lecture Notes in Artificial Intelligence 2650, pages 253–268. Springer Verlag, Berlin, 2003.
14. P. McBurney and S. Parsons. Representing epistemic uncertainty by means of dialectical argumentation. *Annals of Mathematics and Artificial Intelligence*, 32(1–4):125–169, 2001.
15. 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.

16. 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.
17. 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.
18. S. Parsons and P. McBurney. Argumentation-based communication between agents. In M.-P. Huget, editor, *Communication in Multi-Agent Systems: Agent Communication Languages and Conversation Policies*, pages 164–178. Springer Verlag, Berlin, 2003.
19. 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.
20. S. Parsons, M. Wooldridge, and L. Amgoud. An analysis of formal inter-agent dialogues. In *1st International Conference on Autonomous Agents and Multi-Agent Systems*. ACM Press, 2002.
21. S. Parsons, M. Wooldridge, and L. Amgoud. On the outcomes of formal inter-agent dialogues. In *2nd International Conference on Autonomous Agents and Multi-Agent Systems*. ACM Press, 2003.
22. S. Parsons, M. Wooldridge, and L. Amgoud. Properties and complexity of formal inter-agent dialogues. *Journal of Logic and Computation*, 13(3):347–376, 2003.
23. H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127:187–219, 2001.
24. 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.
25. 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.
26. K. Sycara. Argumentation: Planning other agents' plans. In *Proceedings of the Eleventh Joint Conference on Artificial Intelligence*, pages 517–523, 1989.
27. D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. State University of New York Press, Albany, NY, 1995.