Argumentation-based dialogues for agent co-ordination

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)

Peter McBurney
Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, United Kingdom. (p.j.mcburney@csc.liv.ac.uk)

Abstract. Many techniques for coordinating agents require that the agents communicate, and many of the requisite communications need more than the exchange of a few terse illocutions. In other words they require some form of dialogue. This paper discusses one way to create such dialogues, the use of argumentation, and illustrates the use of this approach in the definition of dialogues about joint plans.

1. Introduction

Many techniques for coordinating agents require that the agents communicate, and many of the requisite communications need more than the exchange of a few terse illocutions. In other words they require some form of dialogue. Now, when we humans engage in any form of dialogue it is natural for us to do so in a somewhat skeptical manner. If someone informs us of a fact that we find surprising, we typically question it. Not in an aggressive way, but what might be described as an inquisitive way. When someone tells us "X is true"—where X can range across statements from "It is raining outside" to "The Dow Jones index will continue falling for the next six months"—we want to know "Where did you read that?", or "What makes you think that?". Typically we want to know the basis on which some conclusion was reached. In fact, this questioning is so ingrained that we often present information with some of the answer to the question we expect it to provoke already attached—"It is raining outside, I got soaked through", "The editorial in today's Guardian suggests that consumer confidence in the US is so low that the Dow Jones index will continue falling for the next six months." This is exactly argumentation-based communication. It is increasingly being applied to the design of agent communications languages and frameworks, for example: Dignum and colleagues [12, 13]; Grosz and Kraus [20]; Parsons and Jennings [36, 37]; Reed [45]; Schroeder *et al.* [48]; and Sycara [53]. Indeed, the idea that it is useful for agents to explain what they are doing is not just confined to research on argumentation-based communication [47].

Apart from its naturalness, there are two major advantages of this approach to agent communication. One is that it ensures that agents are *rational* in a certain sense. As we shall see, and as is argued at length in [30], argumentation-based communication allows us to define a form of rationality in which agents only accept statements which they are unable to refute (the exact form of refutation depending on the particular formal properties of the argumentation system they use). In other words agents will only accept things if they don't have a good reason not to. The second advantage builds on this and, as discussed in more detail in [5], provides a way of giving agent communications a *social semantics* in the sense of Singh [51, 52]. The essence of a social semantics is that agents state publicly their beliefs and intentions at the outset of a dialogue, so that future utterances and actions may be judged for consistency against these statements. The truth of an agent's expressions of its private beliefs or intentions can never be fully verified [59], but at least an agent's consistency can be assessed, and, with an argumentation-based dialogue system, the reasons supporting these expressions can be sought. Moreover, these reasons may be accepted or rejected, and possibly challenged and argued-against, by other agents.

The aim of this paper is to sketch the state of the art in argumentation-based agent communication. We will do this not by describing all the relevant work in detail, but by identifying what we consider to be the main issues in building systems that communicate in this way, by briefly describing how our previous work has addressed these, and by giving references to the relevant work and that of other authors. Where such issues have not yet been considered by anyone we will suggest ways in which they could be addressed, but this paper is much more of a survey than a report on new work.

The paper starts in Section 2 by describing some influential work in philosophy to which we refer throughout the paper. Section 3 describes at a high level a number of ways in which argumentation can be used within agent communication, and Section 4 then makes this discussion concrete with a specific example of a system for argumentation-based communication. This system has some new features, but is a minor variant on systems we have discussed before, the modifications being to enable us to capture the kind of reasoning from [36] in a system like [4] for the first time. Section 5 shows this system in action, Section 6 discusses a broad range of relevant work on argumentation, and Section 7 concludes.

2. Philosophical background

Our work on argumentation-based dialogue has been influenced by a model of human dialogues due to argumentation theorists Doug Walton and Erik Krabbe [57]. Walton and Krabbe set out to analyze the concept of commitment in dialogue, so as to "provide conceptual tools for the theory of argumentation" [57, 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 [57]):

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(s) 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 in order to decide what course of action to take in some situation. Participants share

[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.

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.

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.

This framework can be used in a number of ways. First, we have increasingly used this typology as a framework within which it is possible to compare and contrast different systems for argumentation. For example, in [4] we used the classification, and the description of the start conditions and aims of participants given in [57], to show that the argumentation system described in [4] could handle persuasion, information seeking and inquiry dialogues. Second, we have also used the typology as a means of classifying particular argumentation systems. Thus, for example, we can identify the system discussed in [36] as including elements of deliberation (it is about joint action) and persuasion (one agent is attempting to persuade the other to do something different) rather than negotiation as it was originally billed. Similarly the work of Dignum and colleagues [12, 13] is described as deliberation, and is certainly concerned with team building (which has the right focus), but on examination seems to be more accurately described as a deliberation/persuasion hybrid. The same is true of [34] which is also described by the authors as deliberation.

Third, we can use the typology as a means of distinguishing the focus (and thus the detailed requirements for) systems intended to be used for engaging in certain types of dialogue. Thus, for instance, we have defined locutions that can together be used to perform inquiry [31] and deliberation [22] dialogues.

The final aspect of this work that is relevant, in our view, is that it stresses the importance of being able to handle mixed dialogues—for example dialogues of one kind which include embedded dialogues of another kind. Thus, for example, a negotiation dialogue about the purchase of a car might include an embedded information seeking dialogue (to find the buyer's requirements), and an embedded persuasion dialogue (about the value of a particular model). This has led to two proposals for formalism in which dialogues can be combined in this way [32, 45].

3. Argumentation and dialogue

The focus of attention by philosophers to argumentation has been on understanding and guiding human reasoning and argument. It is not surprising, therefore, that this work says little about how argumentation may be applied to the design of communications systems for artificial agents. In this section we consider some of the issues relevant to such application.

3.1. Languages and argumentation

Considering two agents that are engaged in some dialogue, we can distinguish between three different languages that they use. These distinctions are essentially those drawn in [49], although the description of the languages differs, and we have borrowed the same notation for the languages. [2] Each agent has a *base language* that it uses as a means of knowledge representation, a language we might call L. This language can be unique to the agent, or may be the same for both agents. This is the language in which the designer of the agent provides the agent with its knowledge of the world, and it is the language in which the agent's beliefs, desires and intentions (or indeed any other mental notions with which the agent is equipped) are expressed. Given the broad scope of L, it may in practice be a set of languages—for example separate languages for handling beliefs, desires, and intentions—but since all such languages carry out the same function we will regard them as one for the purposes of this discussion.

Each agent is also equipped with a *meta-language* ML. The meta-language, as its name suggests, is a language which expresses facts about another language. In this case the "other language" is the base language L. Agents need meta-languages because, amongst other things, they need to represent their preferences about elements of L. Again ML may in fact be a set of meta-languages and both agents can use different meta-languages. Furthermore, if the agent has no need to make statements about formulae of L, then it may have no meta-language (or, equivalently, it may have a meta-language which it does not make use of). If an agent does have a separate meta-language, then it, like L, is *internal* to the agent.

Finally, for dialogues, the agents need a shared communication language (or two languages such that it is possible to seamlessly translate between them). We will call this language CL. We can consider CL to be a "wrapper" around statements in L and ML, as is the case

[2] Other distinctions are, of course, possible. For the moment those we are using are sufficient for our purposes.

for KQML [17] and the FIPA ACL [18], or a dedicated language into which and from which statements in L or CL are translated. CL might even be L or ML, though, as with ML, we can consider it to be a conceptually different language. The difference, of course, is that CL is in some sense *external* to the agents—it is used to communicate between them. We can imagine an agent reasoning using L and ML, then constructing messages in CL and posting them off to the other agent. When a reply arrives in CL, it is turned into statements in L and ML and these are used in new reasoning.

Argumentation can fit in with these languages in a number of places. First of all, it can be used in internal argumentation—agents can use argumentation as a means of performing their own internal reasoning (as, for example, suggested by Dung [14] and widely studied in AI—see Section 6 for a discussion of this line of work). Similarly, agents can use argumentation for reasoning using ML (which is effectively what [11] does). Independently of whether argumentation is used internally, it can also be used externally, in the sense of being used in conjunction with CL—this is the sense in which Walton and Krabbe [57] consider the use of argumentation in human dialogue and is much closer to the topic of this paper.

3.2. Inter-agent argumentation

External argumentation can happen in a number of ways. The main issue, the fact that makes it argumentation, is that the agents do not just exchange facts but also exchange additional information such as reasons for the facts. In persuasion dialogues, which are by far the most studied type of argumentation-based dialogues, these reasons are typically the reasons why the facts are thought to be true. Thus, if agent A wants to persuade agent B that p is true, it does not just state the fact that p, but also gives, for example, a proof of p based on information (grounds) that A believes to be true. If the proof is sound then B can only disagree with p if either it disputes the truth of some of the grounds or if it has an alternative proof that p is false. The intuition behind the use of argumentation here is that a dialogue about the truth of a claim p moves to a dialogue about the supporting evidence or one about apparently-conflicting proofs. From the perspective of building argumentative agents, the focus is now on how we can bring about either of these kinds of discussion.

There are a number of aspects, in particular, that we need to focus on. These include:

– Clearly communication will be carried out in CL, but it is not clear how arguments will be passed in CL. Will arguments form

separate locutions, or will they be included in the same kind of CL locution as every other piece of information passed between the agents?

- Clearly the exchange of arguments between agents will be subject to some protocol, but it is not clear how this is related, if at all, to the protocol used for the exchange of other messages. Do they use the same protocol? If the protocols are different, how do agents know when to move from one protocol to another?

- Clearly the arguments that agents make should be related to what they know, but it is not clear how best this might be done. Should an agent only be able to argue what it believes to be true? If not, what arguments is an agent allowed to make?

One approach to constructing argumentation-based agents is the way suggested in [49]. In this work CL contains two sets of illocutions. One set allows the communication of facts (in this case statements in ML that take the form of conjunctions of value/attribute pairs, intended as offers in a negotiation). The other set allows the expressions of arguments. These arguments are unrelated to the offers, but express reasons why the offers should be acceptable, appealing to a rich representation of the agent and its environment: the kinds of argument suggested in [49] are threats such as, "If you don't accept this I will tell your boss," promises like: "If you accept my offer I'll bring you repeat business," and appeals such as: "You should accept this because that is the deal we made before."

There is no doubt that this model of argumentation has a good deal of similarity with the kind of argumentation we engage in on a daily basis. However, it makes considerable demands on any implementation. For a start, agents which desire to argue in this manner need very rich representations of each other and their environments (especially compared with agents which simply wish to debate the truth of a proposition given what is in their knowledge-base). Such agents also require an answer to the second two points raised above, and the very richness of the model makes it hard (at least for the authors) to see how the third point can be addressed.

Now, the complicating factor in both of the bullet points raised above is the need to handle two types of information—those that are argument-based and those that aren't. One way to simplify the situation is to make all communication argument-based, and that is the approach that we have been following of late. In fact, we go a bit further than even this suggests, by considering agents that use argumentation both for internal reasoning and as a means of relating what they believe

and what they communicate. We describe this approach in the next section.

3.3. Argumentation at all levels

In more detail what we are proposing is the following. First of all, every agent carries out internal argumentation using L. This allows it to resolve any inconsistency in its knowledge base (which is important when dealing with information from many sources since such information is typically inconsistent) and to establish some notion of what it believes to be true (though this notion is defeasible since new information may come to light that provides a more compelling argument against some fact than there previously was for that fact). The upshot of this use of argumentation, however it is implemented, is that every agent can not only identify the facts it believes to be true but can supply a rationale for believing them.

This feature then provides us with a way of ensuring a kind of rationality of the agents—rationality in communication. It is natural that an agent which resolves inconsistencies in what it knows about the world uses the same technique to resolve inconsistencies between what it knows and what it is told. In other words the agent looks at the reasons for the things it is told and accepts these things provided they are supported by more compelling reasons than there are against the things. If agents are only going to accept things that are backed by arguments, then it makes sense for agents to only say things that are also backed by arguments. Both of us, separately in [30] and [5], have suggested that such an argumentation-based approach is a suitable form of rationality, and it was implicit in [4].[3]

The way that this form of rationality is formalized is, for example, to only permit agents to make assertions that are backed by some form of argument, and to only accept assertions that are so backed. In order words, the formation of arguments becomes a precondition of the locutions of the communication language CL, and the locutions are linked to the agents' knowledge bases.

Although it is not immediately obvious, this gives argumentation-based approaches a *social semantics* in the sense of Singh [51, 52]. The naive reason for this is that since agents can only assert things that in their considered view are true (which is another way of putting the fact that the agents have more compelling reasons for thinking something is true than for thinking it is false), other agents have some guarantee that they are true. However agents may lie, and a suitably

[3] This meaning of rationality is also consistent with that commonly given in philosophy, see, e.g., [24].

sophisticated agent will always be able to simulate truth-telling. A more sophisticated reason is that, assuming such locutions are built into CL, the agent on the receiving end of the assertion can always challenge statements, requiring that the reasons for them are stated. These reasons can be checked against what that agent knows, with the result that the agent will only accept things that it has no reason to doubt. This ability to question statements gives argumentation-based communication languages a degree of verifiability that other semantics, such as the original modal semantics for the FIPA ACL [18], lack.

3.4. Dialogue games

Dialogues may be viewed as games between the participants, called *dialogue games* [25]. In this view, explained in greater detail in [33], each participant is a player with an objective they are trying to achieve and some finite set of moves that they might make. Just as in any game, there are rules about which player is allowed to make which move at any point in the game, and there are rules for starting and ending the game.

As a brief example, consider a persuasion dialogue. We can think of this as being captured by a game in which one player initially believes p to be true and tries to convince another player, who initially believes that p is false, of that fact. The game might start with the first player stating the reason why she believes that p is true, and the other player might be bound to either accept that this reason is true (if she can find no fault with it) or to respond with the reason she believes it to be false. The first player is then bound by the same rules as the second was—to find a reason why this second reason is false or to accept it—and the game continues until one of the players is forced to accept the most recent reason given and thus to concede the game.

This is exactly the form of the dialogue game developed in [4], and, as described in [33], there are a large number of dialogue game formulations of inter-agent dialogues. The approach described in the next section is another, though in theory it is also possible to formulate the same kind of system without making it a dialogue game.

4. A system for argumentation-based communication

In this section we give a concrete instantiation of the rather terse description given in Section 3.3, providing an example of a system for carrying out argumentation-based communication of the kind first suggested in [36]. Our main aim is to illustrate the points made above,

but we also make a minor technical advance in providing an extension of the kind of system we introduced in [4] to handle the kind of reasoning in [36]. [4]

4.1. A system for internal argumentation

We start with a system for internal argumentation—this is an extended version of [14], where the extension allows for a notion of the strength of an argument [3], which is augmented to handle beliefs and intentions. To define this system we start with a propositional language which we call $\mathcal{L}$. From $\mathcal{L}$ we then construct formulae such as $B_i(p)$, $D_i(p)$ and $I_j(q)$ for any p and q which are formulae of $\mathcal{L}$. This extended propositional language, and the compound formulae that may be built from it using the usual logical connectives, is the base language L of the argumentation-based dialogue system we are describing. $B_i(\cdot)$ denotes a belief of agent i, $D_i(\cdot)$ denotes a desire of agent i, and $I_j(\cdot)$ denotes an intention of agent j, so the overall effect of this language is just to force every formula to be a belief, a desire, or an intention. We will denote formulae of L by ϕ, ψ, σ .... Since we are only interested in syntactic manipulation of beliefs, desires and intentions here, we will give no semantics for formulae such as $B_i(p)$ and $B_i(p) \rightarrow D_i(p)$— suitable ways of dealing with the semantics are given elsewhere (e.g. [37, 58]). An agent has a knowledge base Σ which is allowed to be inconsistent, and has no deductive closure. The symbol $\vdash$ denotes classical inference and $\equiv$ denotes logical equivalence.

An argument is a formula of L 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 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).

[4] The minor advance includes introducing a formal protocol for the dialogue in [36], something that was missing from the original, and starting to extend the kind of argumentation in [4] to work with a more complex language than just propositional logic. Still, there is not much new here.

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 notions of rebutting, undercutting and attacking.

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

DEFINITION 3. *Let A_1 and A_2 be two distinct arguments of $\mathcal{A}(\Sigma)$. A_1* rebuts *A_2 iff $Conclusion(A_1)$* attacks *$Conclusion(A_2)$.*

These are the usual notions of rebut and undercut from the AI literature (for example in [14, 16]). For the particular situation we are dealing with here, we need the following notion of "attack":

DEFINITION 4. *Given two distinct formulae h and g of $\mathcal{L}$ such that $h \equiv \neg g$, then, for any i and j:*

- $B_i(h)$ attacks $B_j(g)$*;*
- $D_i(h)$ attacks $D_j(g)$*; and*
- $I_i(h)$ attacks $I_j(g)$.

Note that this notion of attack is a generalization of that in [3], and, while related to that in [37] both extends it (in allowing "attacks" between things other than intentions) and is less extensive than it (by not allowing "attacks" between second order intentions). [5] The differences are determined by the kind of reasoning we are trying to capture. In the case we are dealing with here, it is important to be able to identify conflicts between the propositions within modalities, since the conflict between suggestions made by two agents can be grounded in the fact that one agent believes p and another believes $\neg p$.

With these definitions, an argument is rebutted if it has a conclusion $B_i(p)$ and there is another argument which has as its conclusion $B_j(\neg p)$ or $B_j(q)$ such that $q \equiv \neg p$. An argument with a desire as its conclusion can similarly be rebutted by another argument with a desire as its conclusion, and the same thing holds for intentions. Thus we recognize "Peter intends that this paper be written by the deadline" and "Simon intends this paper not to be written by the deadline" as rebutting each other, along with "Peter believes God exists" and "Simon does not believe God exists", but we do not recognize "Peter intends that

[5] Indeed the language we are using here does not allow the statement of such intentions—they are not necessary for what we wish to do and are therefore omitted.

this paper will be written by the deadline" and "Simon does not believe that this paper will be written by the deadline" as rebutting each other. Undercutting occurs in exactly the same situations, except that it holds between the conclusions of one argument and an element of the support of the other.[6]

For some languages L, and some definitions of "attacks", there is a strong relationship between rebuts and undercuts. Consider, for example, the argumentation system described in [4]. This uses classical propositional logic and defines "attack" as holding between two propositional formulae h and g iff $h \equiv \neg g$. Now, if we have an argument (S, b), then an argument that rebuts (S, b) will always also undercut it[7], and so there is little point in defining rebuttal. Note that the use of $h \equiv \neg g$ within the definition of "attacks" means that $h = Block\ A\ is\ red$ does not attack $g = Block\ A\ is\ green$. In order for an agent to detect a conflict, it would also have to know that $Block\ A\ is\ red \rightarrow \neg Block\ A\ is\ green$, and then this latter could be used along with h to construct a argument that rebutted the argument $(\{g\}, g)$.

To capture the fact that some facts are more strongly believed and intended than others, we assume that any set of facts has a preference order over it[8]. 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 5. *Let A_1 and A_2 be two arguments in $\mathcal{A}(\Sigma)$. A_1 is* preferred *to A_2 according to Pref iff*

$$level(Support(A_1)) \leq level(Support(A_2))$$

[6] Note that attacking and rebutting are symmetric but not reflexive or transitive, while undercutting is neither symmetric, reflexive nor transitive.

[7] As a quick demonstration of why this is the case, consider a proposition a which is part of S. Consider $S - \{a\}$. Since when we add a to it we get, by definition, b, we can use the deduction theorem to add $a \rightarrow b$, where $\rightarrow$ indicates material implication. Since the rebutting argument gives us $\neg b$, we can use modus tollens to give us an argument which attacks a.

[8] We ignore for now the fact that we might require different preference orders over beliefs and intentions and indeed that different agents will almost certainly have different preference orders, noting that the problem of handling a number of different preference orders was considered in [6] and [8].

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

DEFINITION 6. *An* argumentation system *(AS) is a triple*

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

such that:

- $\mathcal{A}(\Sigma)$ *is a set of the arguments built from* Σ,
- *Undercut/Rebut is a binary relation capturing the existence of an undercut or rebut holding between arguments,* $\mathit{Undercut/Rebut} \subseteq \mathcal{A}(\Sigma) \times \mathcal{A}(\Sigma)$, *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 7. *Let* A_1, A_2 *be two arguments of* $\mathcal{A}(\Sigma)$.

- *If* A_2 *undercuts or rebuts* 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 *such that* B *undercuts or rebuts* A *and* A *does not defend itself against* B *then* $\exists$ $C \in \mathcal{S}$ *such that* C *undercuts or rebuts* B *and* B *does not defend itself against* C.

Henceforth, $C_{\mathit{Undercut/Rebut,Pref}}$ will gather all non-undercut and non-rebut arguments along with arguments defending themselves against all their undercutting and rebutting arguments. [2] showed that the set $\underline{\mathcal{S}}$ of acceptable arguments of the argumentation system $\langle \mathcal{A}(\Sigma), \mathit{Undercut/Rebut}, \mathit{Pref} \rangle$ is the least fixpoint of a function $\mathcal{F}$:

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

where $\mathcal{S} \subseteq \mathcal{A}(\Sigma)$.

DEFINITION 8. *The set of* acceptable *arguments of an argumentation system* $\langle \mathcal{A}(\Sigma), \mathit{Undercut}, \mathit{Pref} \rangle$ *is:*

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

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

Note that while we have given a language L for this system, we have given no language ML. This particular system does not have a meta-language (and the notion of preferences it uses is not expressed in a meta-language). It is, of course, possible to add a meta-language to this system—for example, in [6] we added a meta-language which allowed us to express preferences over elements of L, thus making it possible to exchange (and indeed argue about, though this was not done in [6]) preferences between formulae.

4.2. Arguments between agents

Now, this system is sufficient for internal argumentation within a single agent, and the agent can use it to, for example, perform nonmonotonic reasoning and to deal with inconsistent information. To allow for dialogues, we have to introduce some more machinery. Clearly part of this will be the communication language, but we need to introduce some additional elements first. These elements are datastructures which our system inherits from its dialogue game ancestors as well as previous presentations of this kind of system [4, 7].

Dialogues are assumed to take place between two agents, P and C.[9] Each agent has a knowledge base, Σ_P and Σ_C respectively, containing their beliefs. In addition, following Hamblin [21], 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 [7] 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 to both commitment stores. Thus P can make use of

$$\langle \mathcal{A}(\Sigma_P \cup CS(C)), \mathit{Undercut}/\mathit{Rebut}, \mathit{Pref} \rangle^{10}$$

and C can make use of

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

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

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

All the knowledge bases contain propositional formulae and are not closed under deduction, and all 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 [6] 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 [35].

With this background, we can present the set of dialogue moves that we will use, the set which comprises the locutions of CL. For each move, we give what we call rationality rules, dialogue rules, and update rules. These locutions are those from [38] and are based on the rules suggested by [29] which, in turn, were based on those in the dialogue game DC introduced by MacKenzie [28]. The rationality rules specify the preconditions for making the move. Unlike those in [4, 7], these rules are not absolute, but are defined in terms of the agent attitudes discussed below, and these provide the social semantics for the locutions. The update rules specify how commitment stores are modified by the move.

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

assert(ϕ) where ϕ is a formula of L.

rationality: the usual assertion condition for the agent.

update: $CS_i(P) = CS_{i-1}(P) \cup \{\phi\}$ and $CS_i(C) = CS_{i-1}(C)$

Here ϕ can be any formula of L, as well as the special character $\mathcal{U}$, discussed in the next sub-section.

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

rationality: the usual assertion condition for the agent.

update: $CS_i(P) = CS_{i-1} \cup S$ and $CS_i(C) = CS_{i-1}(C)$

The counterpart of these moves are the acceptance moves:

accept(ϕ) ϕ is a formula of L.

rationality: The usual acceptance condition for the agent.

update: $CS_i(P) = CS_{i-1}(P) \cup \{\phi\}$ and $CS_i(C) = CS_{i-1}(C)$

accept(S) S is a set of formulae of L.

rationality: the usual acceptance condition for every $\sigma \in S$.

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(ϕ) where ϕ is a formula of L.

rationality: $\emptyset$

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

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

question(ϕ) where ϕ is a formula of L.

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 $\mathcal{M}^d_{\mathcal{DC}}$. These locutions are the bare minimum to carry out a dialogue, and, as we will see below, require a fairly rigid protocol with a lot of aspects implicit. Further locutions such as those discussed in [32], would be required to be able to debate the beginning and end of dialogues or to have an explicit representation of movement between embedded dialogues.

The locutions in $\mathcal{M}^d_{\mathcal{DC}}$ say nothing about how the preferences of an agent are updated. This is intentional. Here we assume that the preferences do not change as a result of locutions—we assume that an agent already has a preference order over all possible formulae at the start of a dialogue so that any new formula it *accept*s just slots into the existing order. Clearly this is a gross simplification, adopted here to shorten the presentation. It is easy enough, as in [6], to add in a language ML which explicitly states an agent's preferences, and to allow locutions to be made about these preferences and agents to be persuaded to change their preferences. Note that these locutions are essentially those of $\mathcal{M}'_{\mathcal{DC}}$ [38], modified to deal with the slightly more complex base language we have here. We have previously shown that these allow us to handle information seeking, inquiry and persuasion dialogues from the Walton and Krabbe classification. Here we use them to carry out a form of deliberation.

Now, the set of moves/locutions $\mathcal{M}^d_{\mathcal{DC}}$ defines the communication language CL, and hopefully it is reasonably clear from the description so far how argumentation between agents takes place; a prototypical persuasion dialogue is as follows:

1. P has an acceptable argument $(S, B_P(p))$, built from Σ_P, and wants C to accept $B_P(p)$. Thus, P asserts $B_P(p)$.

2. C has an argument $(S', B_C(\neg p))$ and so cannot accept $B_P(p)$. Thus, C asserts $B_C(\neg p)$.

3. P cannot accept $B_C(\neg p)$ and challenges it.

4. C responds by asserting S'.

5. P has an argument $(S'', B_P(\neg q))$ where $B_C(q) \in S'$, and asserts $B_P(\neg q)$.

6. C challenges $B_P(\neg q)$.

7. ...

At each stage in the dialogue agents can build arguments using information from their own private knowledge base, and the propositions made public (by assertion into commitment stores).

4.3. Rationality and protocol

The final part of the abstract model we introduced above was the use of argumentation to relate what an agent "knows" (in this case what is in its knowledge-base and the commitment stores) and what it is allowed to "say" (in terms of which locutions from CL it is allowed to utter). We make this connection by specifying the rationality conditions in the definitions of the locutions and relating these to what arguments an agent can make. We do this as follows, essentially defining different types of rationality [38].

DEFINITION 9. *An agent may have one of three* assertion *attitudes.*

- *a* confident *agent can assert any formula ϕ for which there is an argument (S, ϕ).*
- *a* careful *agent can assert any formula ϕ for which there is an argument (S, ϕ) if no stronger rebutting argument exists.*
- *a* thoughtful *agent can assert any formula ϕ for which there is an acceptable argument (S, ϕ).*

Thus a thoughtful agent will only put forward formulae which, so far as it knows, are correct. A careful agent will only put forward formulae which aren't directly rebutted. A confident agent won't stop to make either of these checks.[11]

[11] Note that, as a first step, we define these agent attributes uniformly; in later work, we will consider agents which assert or accept formulae in a context-dependent manner.

Of course, defining when an agent can assert formulae is only one half of what is needed. The other part is to define the conditions on agents accepting formulae. Here we have the following [38].

DEFINITION 10. *An agent may have one of three* acceptance *attitudes.*

- *a* credulous *agent can accept any formula ϕ for which there is an argument (S, ϕ).*
- *a* cautious *agent can accept any formula ϕ for which there is an argument (S, ϕ) if no stronger rebutting argument exists.*
- *a* skeptical *agent can accept any formula ϕ for which there is an acceptable argument (S, ϕ).*

In order to complete the definition of the system, we need only to give the protocol that specifies how a dialogue proceeds. This we do below, providing a protocol (which was not given in the original) for the kind of example dialogue given in [36, 37]. As in those papers, the kind of dialogue we are interested in here is a dialogue about joint plans, and in order to describe the dialogue, we need an idea of what one of these plans looks like:

DEFINITION 11. *An* plan *is an argument $(S, I_i(p))$. $I_i(p)$ is known as the* subject *of the plan.*

Thus a plan is just an argument for a proposition that is intended by some agent. The detail of "acceptable" and "attack" ensure that an agent will only be able to assert or accept a plan if there is no intention which is preferred to the subject of the plan so far as that agent is aware (given the checks it carries out given its attitude), and there is no conflict between any elements of the support of the plan. We then have the following protocol, which we will call $\mathcal{D}$ for a dialogue between agents A and B.

1. If allowed by its assertion attitude, A asserts both the conclusion and support of a plan $(S, I_A(p))$. If A cannot assert any $I_A(p)$, the dialogue ends.
2. B *accepts* $I_A(p)$ and S if possible. If both are accepted, the dialogue terminates.
3. If the $I_A(p)$ and S are not accepted, then B *asserts* the conclusion and support of an argument (S', ϕ) which undercuts or rebuts $(S, I_A(p))$.

4. *A asserts* either the conclusion and support of $(S''', I_A(p))$, which does not undercut or rebut (S', ϕ), or the statement $\mathcal{U}$. In the first case, the dialogue returns to Step 2; in the second case, the dialogue terminates.

The utterance of a statement $\mathcal{U}$ indicates that an agent is unable to add anything to the dialogue, and so the dialogue terminates whenever either agent asserts this.

Note that in B's response it need not assert a plan (A is the only agent which has to mention plans). This allows B to disagree with A on matters such as the resources assumed by A ("No, I don't have the car that week"), or the tradeoff that A is proposing ("I don't want your Megatokyo T-shirt, I have one like that already"), even if they don't directly affect the plans that B has.

As it stands, the protocol is a rather minimalist but suffices to capture the kind of interaction in [36, 37]. One agent makes a suggestion which suits it (and may involve the other agent). The second looks to see if the plan prevents it achieving any of its intentions, and if so has to put forward an argument which clashes in some way (we could easily extend the protocol so that B does not have to put forward this argument, but can instead engage A in a persuasion dialogue about A's plan in a way that was not considered in [36, 37]). The first agent then has the chance to respond by either finding a non-clashing way of achieving what it wants to do or suggesting a way for the second agent to achieve its intention (if one is mentioned) without clashing with the first agent's original plan.

There is also much that is implicit in the protocol, for example: that the agents have previously agreed to carry out this kind of dialogue (since no preamble is required); that the agents are basically co-operative (since they accept suggestions if possible); and that they will end the dialogue as soon as a possible agreement is found or it is clear that no progress can be made (so neither agent will try to filibuster for its own advantage). Such assumptions are consistent with Grice's co-operative maxims for human conversation [19].

One advantage of such a minimal protocol is that it is easy to show that the resulting dialogues have some desirable properties. The first of these is that the dialogues terminate:

PROPOSITION 1. *A dialogue under protocol $\mathcal{D}$ between two agents G and H with any acceptance and assertion attitudes will terminate.*

Proof: *$\mathcal{D}$ requires that one agent asserts the conclusion and support of an argument, and this is either accepted or the agent asserts another pair of conclusion and support, and this is either accepted or the agent asserts another pair of conclusion and support argument, and so on.*

If one of these pairs is accepted the dialogue terminates, and if one agent utters the same argument twice the dialogue terminates. Since the agents' knowledge is finite, there are a finite number of arguments that can be uttered before the dialogue terminates, and so the dialogue will always terminate. □

If both agents are thoughtful and skeptical, we can also obtain conditions on the result of the dialogue:

PROPOSITION 2. *Consider a dialogue under protocol $\mathcal{D}$ between two thoughtful/skeptical agents G and H, where G starts by uttering a plan with the subject $I_G(p)$.*

- *If the dialogue terminates with the utterance of $\mathcal{U}$, then there is no plan with the subject $I_G(p)$ in $A(\Sigma_G \cup CS(H))$ that H can accept.*

- *If the dialogue terminates without the utterance of $\mathcal{U}$, then there is a plan with the subject $I_G(p)$ in $A(\Sigma_G \cup \Sigma_H)$ that is acceptable to both G and H.*

Proof: *The proof follows almost directly from the protocol. Let's start by considering the second part of the theorem. G starts by asserting the conclusion and support of a plan $(S, I_G(p))$ it finds acceptable (since this is all a thoughtful agent can assert). If H finds both parts acceptable it will accept it (note that "acceptable" and "accept" only coincide like this for skeptical agents), satisfying the theorem. If H does not find both parts of $(S, I_G(p))$ acceptable, it does not accept the offending part, and by the definition of a thoughtful agent, this is because it has a rebutting or undercutting argument, and its response is to assert this argument. A then has to respond with the conclusion and support of another plan, and the theorem is again validated if H accepts it.*

Now, given the finiteness of the agents' knowledge, unless H accepts a conclusion/support pair, making the second part of the theorem true, eventually G will be in step 4 of the protocol and utter the statement $\mathcal{U}$. At that point the dialogue will terminate and there are no plans that can be constructed from the knowledge that G has (its own knowledge and those things that H has stated) which are acceptable to both it and H that have the subject $I_G(p)$. □

Note that since we can't determine exactly what H says, and therefore what the contents of $CS(H)$ are, we are not able to make the two parts of the theorem symmetrical (or the second part an "if and only if", which would be the same thing).

Thus if the agents reach agreement, it is an agreement on a plan which neither of them has any reason to think problematic. In [36, 37] we called this kind of dialogue a negotiation. From the perspective of Walton and Krabbe's typology it isn't a negotiation—it is closer to a deliberation with the agents discussing what they will do. However, it seems to be rather asymmetric when compared with what Walton and Krabbe had in mind for a deliberation (and is certainly more limited than the kind of deliberation dialogues we discuss in [22]). First, only one agent gets to suggest plans for both to consider (B does not really make suggestions, just points out why A's suggestions don't work), and second, the plans are presented as monolithic entities rather than being constructed in discussion (when A makes a suggestion, it is a suggestion for a complete plan, so that B is only able to "take it or leave it" rather than make modifications or suggestions). We could, of course, easily devise a less asymmetric kind of protocol where both agents were allowed to suggest entire plans, or one in which plans are constructed step by step rather like the inquiry dialogues in [38].

Finally, we should note another limitation of the protocol. Because the protocol insists that agents consider the conclusion and support of plans together (which is necessary if we are going to produce dialogues like that in [36]), then the dialogue may well fail in the following way. A proposes a plan, B gives a counter-argument, and A cannot produce an alternative plan. If A could produce a counter-argument to B's first argument, then the dialogue might be considered to be failing when it should succeed with the acceptance of A's plan. If we depart from the procedure assumed in [36] then we can solve this problem—indeed we can allow A and B to counter-argue against anything the other says easily enough, allowing much more flexible interactions (though one might consider that this, essentially a mixture of persuasion and deliberation, might be better handled by having persuasion dialogues nested inside a deliberation dialogue much like the current one).

5. An example

In this section we show how the system given in the previous section can handle the nail example from [36]. The example concerns a home improvement agent which has the intentions of doing some work around a house. This agent, Agent 1, has the intention of hanging a picture, and knows that it has in its possession a picture, a hammer, and a nail. It also believes that once it has a picture, a hammer and a nail, then it has all it needs to go about hanging a picture, and it has some general information to the effect that if an agent can do something, and intends

to do that something, then it should go ahead and do it.

$$
\begin{array}{ll}
D_1(Do(agent_1, hang\_picture)) & f1 \\
B_1(Have(agent_1, picture)) & f2 \\
B_1(Have(agent_1, nail)) & f3 \\
B_1(Have(agent_1, hammer)) & f4 \\
B_1(Have(W, hammer)) \wedge B_1(Have(X, nail)) & \\
\quad \wedge B_1(Have(Y, picture)) \rightarrow B_1(Can(Z, hang\_picture)) & r1 \\
B_1(Can(X, Y)) \wedge D_1(Do(X, Y)) \rightarrow I_1(Do(X, Y)) & r2
\end{array}
$$

Two points need to be made about this knowledge. The first is the use of symbols such as $f1$. These just identify formulae, and allow us to write supports in a compact fashion (unlike in [36] they are not part of the language). The second is that the language we are using here isn't propositional, in contrast to the argumentation system we introduced before. The reason for that is the same as the reason for introducing quantifiers—it allows us to write things more compactly. Given that we don't really use variables (we could re-write $r1$ and $r2$ with the variables instantiated to every possible combination of $agent_1$ and $agent_2$ with no change in the information expressed) and terms like $Have(agent_1, picture)$ is treated as if they are atomic, the language we have here is functionally equivalent to propositional logic. It is just easier to read and write.

This information is broadly that in the example in [36], though it seems to us to be closer to the use made of beliefs, desires and intentions by Bratman *et al.* [10] than that in [36]. In particular, $r2$ captures something like the main function of a BDI interpreter—if an agent is able to do something and desires to do it, then it should adopt the intention of doing it. Of course, translating this into a single logical implication loses something, and the same is true of $r1$. This latter is intended to capture the essence of the plan-building the agent does, and is intentionally simple. Creating a more realistic logic-based planner would detract from the argumentation that is our main focus. [12] Now, from the information it has, Agent 1 can build the following argument:

$$(\{f1, f2, f3, f4, r1, r2\}, I_1(Do(agent_1, hang\_picture)))$$

indicating that it has a plan for hanging the picture.

Now consider the following variation of the example to the case in which there are two home-improvement agents with different objectives

[12] Note, however, that it is intended that $r1$ make it possible for agent 1 to infer that when a has the hammer and b has the nail and c has the picture then d can hang a picture—this seems to us to be appropriate for a simple co-operative planning domain of the kind in this example.

and different resources. Agent 1 is much as described before, however, it now has a screw and a screwdriver rather than a nail, knows how to hang mirrors as well as pictures, and furthermore, knows that Agent 2 has a nail:

$$D_1(Do(agent_1, hang\_picture)) \quad f1$$
$$B_1(Have(agent_1, picture)) \quad f2$$
$$B_1(Have(agent_1, screw)) \quad f3$$
$$B_1(Have(agent_1, hammer)) \quad f4$$
$$B_1(Have(agent_1, screwdriver)) \quad f5$$
$$B_1(Have(agent_2, nail)) \quad f6$$
$$B_1(Have(W, hammer)) \wedge B_1(Have(X, nail)) \wedge B_1(Have(Y, picture)) \rightarrow B_1(Can(Z, hang\_picture)) \quad r1$$
$$B_1(Have(W, screwdriver)) \wedge B_1(Have(X, screw) \wedge B_1(Have(Y, mirror)) \rightarrow B_1(Can(Z, hang\_mirror)) \quad r2$$
$$B_1(Can(X, Y)) \wedge D_1(Do(X, Y)) \rightarrow I_1(Do(X, Y)) \quad r3$$
$$B_2(Have(X, Y)) \rightarrow B_1(Have(X, Y)) \quad r4$$

The final rule here is intended to, rather roughly, handle communication between agents that trust one another. If Agent 2 asserts that it believes one agent has something, then Agent 1 has a prima facie case to believe that as well (it may, of course, be overturned by a stronger argument to the contrary).

Now, Agent 2 knows about hanging mirrors and has the objective of hanging one, but lacks the resources to hang the mirror on its own:

$$D_2(Do(agent_2, hang\_mirror)) \quad f7$$
$$B_2(Have(agent_2, mirror)) \quad f8$$
$$B_2(Have(agent_2, nail)) \quad f9$$
$$B_2(Have(W, hammer)) \wedge B_2(Have(X, nail)) \wedge B_2(Have(Y, mirror)) \rightarrow B_2(Can(Z, hang\_mirror)) \wedge B_2(\neg Have(X, nail)) \quad r5$$
$$B_2(Can(X, Y)) \wedge D_2(Do(X, Y) \rightarrow I_2(Do(X, Y)) \quad r6$$
$$B_1(Have(X, Y)) \rightarrow B_2(Have(X, Y)) \quad r7$$

Agent 1 can work out that it is unable to hang the picture on its own because it is unable to build a plan for $I_1(Do(agent_1, hang\_picture))$ without using Agent 2's nail, but it can build a plan for

$$I_1(Do(agent_1, hang\_picture))$$

that does include the use of the nail:

$$(\{f1, f2, f4, f6, r1, r3, r7\}, I_1(Do(agent_1, hang\_picture)))$$

This argument is acceptable to Agent 1 since it is unable to build any arguments which rebut or undercut it, and it starts the dialogue by asserting it.

Agent 2 then tries to accept the plan. It finds that with the additional information that Agent 2 passes about its resources, it can build a plan for hanging its mirror using Agent 1's hammer:

$$(\{f4, f7, f8, f9, r5, r6, r7\}, I_2(Do(agent_2, hang\_mirror)))$$

and not only that, but it can build the following argument which undercuts Agent 1's plan by attacking $f6$:

$$(\{f4, f7, f8, f9, r5, r6, r7\}, B_2(\neg Have(agent_2, nail)))$$

Agent 2 then passes this latter argument to Agent 1.

Now equipped with the information that Agent 2 has the objective of hanging a mirror, and that this is blocked by the use of its nail to hang Agent 1's picture, Agent 1 can use its mirror-hanging knowledge to propose a different course of action which results in both mirror and picture being hung:

$$(\{f1, f2, f3, f4, f5, f6, f7, f8, r1, r2, r3, r4, r6\},$$
$$I_1(Do(agent_1, hang\_picture)) \wedge I_2(Do(agent_2, hang\_mirror)))$$

Now, this is acceptable to Agent 1, and satisfies the protocol (achieving the subject of Agent 1's original plan as well as providing a plan to achieve Agent 2's goal). Agent 2 then tries to accept this plan, and finds that it can. The dialogue then terminates with success.

6. Other work on argumentation

Argumentation, the study of the process by which agents attempt to convince one another of certain propositions, has been studied in philosophy since at least the time of Aristotle [9]. However, the last five decades have seen a flowering of research on argumentation theory, by philosophers such as Toulmin [55], Lorenzen and Lorenz [26], and Hamblin [21]. Much of this effort has focused on dialectical aspects of argument, for example in the work of Van Eemeren and Grootendorst [15] and of Walton and Krabbe [57]. Following Loui [27], the formal study of argumentation and argumentation has become of great interest to researchers in Artificial Intelligence, particularly in nonmonotonic and uncertain reasoning and in multi-agent systems.

One main approach to the use of argumentation as a technique for nonmonotonic reasoning is the *acceptability* approach. Here the work of Dung [14] has been particularly influential (not least upon the development of the approach we base this work on [1]), and has echoes in the work of Prakken and Sartor [42, 43, 44], Pollock [40], and Vreeswijk

[56]. Although these papers differ in technical detail and the underlying formal language, all of them use notions of undercutting and defeat among arguments to define criteria for acceptability of arguments or propositions supported by arguments.

The main characteristic of the acceptability-based approach to argumentation is that any proposition is considered to hold or not hold depending on the acceptability or otherwise of the argument for it. An alternative, explored by Elvang-Gørannsen *et al.* [16], Pinkas and Loui [39] and Simari and Loui [50] is to classify arguments (and hence their associated propositions) in more detail based upon their relationship with other arguments. Thus, for instance, Elvang-Gørannsen *et al.* identify distinguish arguments that are tautological, unattacked, rebutted and undercut.

All the work described so far was concerned with a single agent reasoning about what to believe. However, it is a small step to considering two or more agents carrying out the kind of procedure by which acceptability is determined—first one agent proposes an argument, another counters with an undercutting or rebutting argument, and so on until one cannot respond and "loses". Exactly this kind of exchange was the concept at the heart of proof-theoretic methods of determining acceptability [1, 44], and moving from the concept to real multi-agent exchanges is simple [4].

The focus or much of the argumentation research in multi-agent systems (for a few representative examples see [37] and [54]) was the application of argumentation for negotiation and reaching agreement. Authors argue that all mechanisms for negotiation have at their heart an exchange of offers. Agents make offers that they find acceptable and respond to offers made to them. Argumentation-based negotiation allows offers to be supported by arguments, which broadly speaking equate to explanations for why the offer was made. This permits greater flexibility than in other negotiation schemes since, for instance, it makes it possible to persuade agents to change their view of an offer by introducing new factors in the middle of a negotiation (just as a car sales person might throw in free insurance to clinch a deal). However, this work (at least until [4]) did not explain when arguments can be used within a negotiation and how they should be dealt with by the agent that receives them, a gap that, as described here, we now believe we have filled.

Despite the focus on negotiation, possibly even more work has been done on persuasion dialogues. This work can be divided into two main groups. The first group, of which [28, 46, 60] are a representative selection, tries to add a reasoning model to a dialogue system in order to handle the different conflicts (inter or intra) agents which may arise

during a dialogue. The second group of persuasion models handle the proof theory of an argumentation system (developed in nonmonotonic reasoning) as a persuasion dialogue between an opponent and a proponent, for example [41], an approach which has much resonance in the AI and Law field.

The remaining kinds of dialogue (in the Walton and Krabbe typology) have been little studied. Hulstijn [23] provides a formulation of inquiry dialogues, as do we elsewhere [31, 38], the latter also discussing information-seeking dialogues, and we have also studied deliberation dialogues [22], but there is no more work that we are aware of.

7. Summary

Argumentation-based approaches to inter-agent communication are becoming more widespread as mechanisms for agent co-ordination, and there are a variety of systems for argumentation-based communication that have been proposed. Many of these address different aspects of the communication problem, and it can be hard to see how they relate to one another. This paper has attempted to put some of this work in context by describing in general terms how argumentation might be used in inter-agent communication, and then illustrating this general model by providing a concrete instantiation of it, finally describing all the aspects required by the example first introduced in [36].

The work that we have described here is still far from complete. Our overall aim is to provide a comprehensive account of inter-agent dialogues, and to build systems capable of supporting such dialogues. There are two main steps that still need to be taken (at least there are two that are immediately obvious to us). One is to extend our analysis to more complex forms of dialogue, such as the deliberation dialogue introduced in [22]. The second is to start building an implementation of this work so that we can experiment with different kinds of dialogue and start to assess what formal dialogue systems are useful in practice.

Acknowledgements
The authors would like to thank Leila Amgoud and Nicolas Maudet for their contribution to the development of many of the parts of the argumentation system described here. We are also very grateful to the reviewers of this paper—their many perceptive comments helped us to greatly improve the paper.

References

1. L. Amgoud. *Contribution à l'intégration des préférences dans le raisonnement argumentatif*. Thèse de doctorat, Université Paul Sabatier, Toulouse, July 1999. (in French).
2. L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation framework. In *Proceedings of the 14th Conference Uncertainty in Artificial Intelligence*, pages 1–7, 1998.
3. L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34:197–215, 2002.
4. L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In E. Durfee, editor, *Proceedings of the 4th International Conference on Multi-Agent Systems*, pages 31–38, Boston, MA, USA, 2000. IEEE Press.
5. 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.
6. 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.
7. L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proceedings of the 14th European Conference on Artificial Intelligence*, pages 338–342, Berlin, Germany, 2000. IOS Press.
8. L. Amgoud, S. Parsons, and L. Perrussel. An argumentation framework based on contextual preferences. In J. Cunningham, editor, *Proceedings of the International Conference on Pure and Applied Practical Reasoning*, London, UK, 2000. Technical Report, Department of Computing, Imperial College, University of London.
9. Aristotle. *Topics*. Clarendon Press, Oxford, UK, 1928. (W. D. Ross, Editor).
10. M. E. Bratman, D. J. Israel, and M. E. Pollack. Plans and resource-bounded practical reasoning. *Computational Intelligence*, 4:349–355, 1988.
11. G. Brewka. Dynamic argument systems: a formal model of argumentation processes based on Situation Calculus. *Journal of Logic and Computation*, 11(2):257–282, 2002.
12. F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Agent theory for team formation by dialogue. In C. Castelfranchi and Y. Lespérance, editors, *Intelligent Agents VII*, pages 141–156, Berlin, Germany, 2001. Springer.
13. F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Creating collective intention through dialogue. *Logic Journal of the IGPL*, 9(2):305–319, 2001.
14. 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.
15. F. H. van Eemeren and R. Grootendorst. *Argumentation, Communication and Fallacies: A Pragma-Dialectical Perspective*. LEA, Mahwah, NJ, USA, 1992.
16. M. Elvang-Gørannson, P. J. Krause, and J. Fox. Dialectic reasoning with inconsistent information. In D. Heckerman and A. Mamdani, editors, *Proceedings of the 9th Conference on Uncertainty in Artifical Intelligence*, pages 114–121, San Mateo, CA, USA, 1993. Morgan Kaufmann.
17. T. Finin, Y. Labrou, and J. Mayfield. KQML as an agent communication language. In J. Bradshaw, editor, *Software Agents*. MIT Press, Cambridge, MA, 1995.

18. FIPA. Communicative Act Library Specification. Technical Report XC00037H, Foundation for Intelligent Physical Agents, 10 August 2001.
19. H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics III: Speech Acts*, pages 41–58. Academic Press, New York City, NY, USA, 1975.
20. 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.
21. C. L. Hamblin. *Fallacies*. Methuen, London, UK, 1970.
22. D. Hitchcock, P. McBurney, and S. Parsons. A framework for deliberation dialogues. In H. V. Hansen, C. W. Tindale, J. A. Blair, and R. H. Johnson, editors, *Proceedings of the 4th Biennial Conference of the Ontario Society for the Study of Argumentation*, Windsor, Ontario, Canada, 2001.
23. J. Hulstijn. *Dialogue models for inquiry and transaction*. PhD thesis, Universiteit Twente, Enschede, The Netherlands, 2000.
24. R. Johnson. *Manifest Rationality: A Pragmatic Theory of Argument*. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2000.
25. J. A. Levin and J. A. Moore. Dialogue-games: metacommunications structures for natural language interaction. *Cognitive Science*, 1(4):395–420, 1978.
26. P. Lorenzen and K. Lorenz. *Dialogische Logik*. Wissenschaftliche Buchgesellschaft, Darmstadt, Germany, 1978.
27. R. Loui. Defeat among arguments: a system of defeasible inference. *Computational Intelligence*, 3:100–106, 1987.
28. J. D. MacKenzie. Question-begging in non-cumulative systems. *Journal of Philosophical Logic*, 8:117–133, 1979.
29. 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.
30. P. McBurney. *Rational Interaction*. PhD thesis, Department of Computer Science, University of Liverpool, 2002.
31. 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*, San Francisco, CA, 2000. Morgan Kaufmann.
32. 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.
33. P. McBurney and S. Parsons. Dialogue game protocols. In Marc-Philippe Huget, editor, *Agent Communications Languages*, Berlin, Germany, 2003. Springer Verlag.
34. P. Panzarasa, N. R. Jennings, and T. J. Norman. Formalizing collaborative decision-making and practical reasoning in multi-agent systems. *Journal of Logic and Computation*, 12(1):55–117, 2002.
35. 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.
36. S. Parsons and N. R. Jennings. Negotiation through argumentation — a preliminary report. In *Proceedings of the 2nd International Conference on Multi-Agent Systems*, pages 267–274, 1996.
37. 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.

38. S. Parsons, M. Wooldridge, and L. Amgoud. An analysis of formal interagent dialogues. In C. Castelfranchi and W. L. Johnson, editors, *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 394–401, New York, USA, 2002. ACM Press.
39. G. Pinkas and R. P. Loui. Reasoning from inconsistency: a taxonomy of principles for resolving conflicts. In *Proceedings of the 3rd International Conference on Principles of Knowledge representation and Reasoning*, pages 709–719, 1992.
40. J. L. Pollock. How to reason defeasibly. *Artificial Intelligence*, 57:1–42, 1992.
41. H. Prakken. On dialogue systems with speech acts, arguments, and counterarguments. In M. Ojeda-Aciego, M. I. P. de Guzman, G. Brewka, and L. M. Pereira, editors, *Proceedings 7th European Workshop on Logic in Artificial Intelligence*, LNAI 1919, pages 224–238, Berlin, Germany, 2000. Springer.
42. H. Prakken. Relating protocols for dynamic dispute with logics for defeasible argumentation. *Synthese*, 127:187–219, 2001.
43. H. Prakken and G. Sartor. Argument-based extended logic programming with defeasible priorties. *Journal of Applied Non-Classical Logics*, 7:25–75, 1997.
44. H. Prakken and G. Sartor. Modelling reasoning with precedents in a formal dialogue game. *Artificial Intelligence and Law*, 6:231–287, 1998.
45. C. Reed. Dialogue frames in agent communications. In Y. Demazeau, editor, *Proceedings of the 3rd International Conference on Multi-Agent Systems*, pages 246–253. IEEE Press, 1998.
46. N. Rescher. *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge*. State University of New York Press, Albany, NY, USA, 1977.
47. P. Riley, P. Stone, and M. Veloso. Layered disclosure: Revealing agents' internals. In C. Castelfranchi and Y. Lespérance, editors, *Intelligent Agents VII*, pages 61–72, Berlin, Germany, 2001. Springer.
48. 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.
49. C. Sierra, N. R. Jennings, P. Noriega, and S. Parsons. A framework for argumentation-based negotiations. In M. P. Singh, A. Rao, and M. J. Wooldridge, editors, *Intelligent Agents IV*, pages 177–192, Berlin, Germany, 1998. Springer.
50. G. R. Simari and R. P. Loui. A mathematical treatment of defeasible reasoning and its implementation. *Artificial Intelligence*, 53:125–157, 1992.
51. M. P. Singh. Agent communication languages: Rethinking the principles. In *IEEE Computer 31*, pages 40–47, 1998.
52. 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.
53. K. Sycara. Argumentation: Planning other agents' plans. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 517–523, 1989.
54. K. Sycara. Persuasive argumentation in negotiation. *Theory and Decision*, 28:203–242, 1990.
55. S. E. Toulmin. *The Uses of Argument*. Cambridge University Press, Cambridge, UK, 1958.
56. G. A. W. Vreeswijk. Abstract argumentation systems. *Artificial Intelligence*, 90:225–279, 1997.
57. D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press, Albany, NY, 1995.

58. M. J. Wooldridge. *Reasoning about Rational Agents*. MIT Press, Cambridge, MA, USA, 2000.
59. M. J. Wooldridge. Semantic issues in the verification of agent communication languages. *Journal of Autonomous Agents and Multi-Agent Systems*, 3(1):9–31, 2000.
60. S. Zabala, I. Lara, and H. Geffner. Beliefs, reasons and moves in a model for argumentation dialogues. In *Proceedings of the Latino-American Conference on Computer Science*, 1999.