

Argumentation-based Communication between Agents

Simon Parsons^{1,2} and Peter McBurney²

¹ 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

² Department of Computer Science
University of Liverpool
Liverpool, L69 7ZF, UK
{s.d.parsons,p.j.mcburney}@csc.liv.ac.uk

Abstract. One approach to agent communication is to insist that agents not only send messages, but support them with reasons why those messages are appropriate. This is argumentation-based communication. This chapter looks at some of our work on argumentation-based communication, focusing on issues which we think apply to all work in this area, and discussing what we feel are the important issues in developing systems for argumentation-based communication between agents.

1 Introduction

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 communication languages and frameworks, for example: Dignum and colleagues [8, 9]; Grosz and Kraus [14]; Parsons and Jennings [25, 26]; Reed [28]; Schroeder *et al.* [30]; and Sycara [34]. 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 [29].

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 [21], argumentation-based communication allows us to define a form of rationality in which agents only accept statements that

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 [4], provides a way of giving agent communications a *social semantics* in the sense of Singh [32, 33]. 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 [37], 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.

This chapter sketches 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, and briefly describing how our work has addressed them.

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 [35]. Walton and Krabbe set out to analyze the concept of commitment in dialogue, so as to "provide conceptual tools for the theory of argumentation" [35, 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. As defined by Walton and Krabbe, the six primary dialogue types 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.

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.¹

Deliberation Dialogues: Participants collaborate to decide what course of action to take in some situation.

Eristic Dialogues: Participants quarrel verbally as a substitute for physical fighting, with each aiming to win the exchange.

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.

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 [3] we used the classification, and the description of the start conditions and aims of participants given in [35] to show that the argumentation system described in [3] could handle persuasion, information seeking and inquiry dialogues. Second, we have also used the classification as a means of classifying particular argumentation systems, for example identifying the system in [25] 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. 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 as in our work to define locutions to perform inquiry [22] and deliberation [16] dialogues.

The final aspect of this work that is relevant, in our view, is that it stresses the importance of being able to handle dialogues of one kind that include embedded dialogues of another kind. Thus 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 formalisms in which dialogues can be combined [23, 28].

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 communication 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. 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 which expresses facts about 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 for KQML [11] or the FIPA ACL [12], 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 be used with these languages in a number of ways. Agents can use argumentation as a means of performing their own internal reasoning either in L , ML , or both. 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 [35] consider the use of argumentation in human dialogue and is much more on the topic of this chapter.

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. 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 [31]. 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 [31] 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 wish 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 [21] and [4], have suggested that such an argumentation-based approach is a suitable form of rationality, and it was implicit in [3].²

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 other 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 [32, 33]. 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 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 [12], lack.

3.4 Dialogue games

Dialogues may be viewed as games between the participants, called *dialogue games* [18]. In this view, explained in greater detail in [24], 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.

² This meaning of rationality is also consistent with that commonly given in philosophy, see, e.g., [17].

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 [25].

4.1 A system for internal argumentation

We start with a system for internal argumentation—this is an extended version of [10], where the extension allows for a notion of the strength of an argument [2], 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 $\phi, \psi, \sigma \dots$. 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. [26, 36]). 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 $H = \text{Support}(A)$ 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 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 $\exists h \in \text{Support}(A_2)$ such that $\text{Conclusion}(A_1)$ *attacks* h .

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

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 $I_j(g)$.

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 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.³

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.⁴ We suppose that this ordering derives from the fact that the knowledge base Σ is stratified into non-overlapping sets $\Sigma_1, \dots, \Sigma_n$ 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 $level(Support(A_1)) \leq level(Support(A_2))$.

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

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

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, $Undercut/Rebut \subseteq \mathcal{A}(\Sigma) \times \mathcal{A}(\Sigma)$, and
- *Pref* is a (partial or complete) preordering on $\mathcal{A}(\Sigma) \times \mathcal{A}(\Sigma)$.

³ Note that attacking and rebutting are symmetric but not reflexive or transitive, while undercutting is neither symmetric, reflexive nor transitive.

⁴ 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 [5] and [7].

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

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

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

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

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

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 [5] 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 [5]) 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 [3, 6].

Dialogues are assumed to take place between two agents, P and C .⁵ Each agent has a knowledge base, Σ_P and Σ_C respectively, containing their beliefs. In addition, following Hamblin [15], each agent has a further knowledge base, accessible to both agents, containing commitments made in the dialogue. These commitment stores are denoted $CS(P)$ and $CS(C)$ respectively, and in this dialogue system (unlike that of [6] for example) an agent's commitment store is just a subset of its knowledge base. Note that the union of the commitment stores can be viewed as the state of the dialogue at a given time. Each agent has access to their own private knowledge base and to both commitment stores. Thus P can make use of $\langle \mathcal{A}(\Sigma_P \cup CS(C)), \text{Undercut/Rebut}, \text{Pref} \rangle$, and C can make use of $\langle \mathcal{A}(\Sigma_C \cup CS(P)), \text{Undercut/Rebut}, \text{Pref} \rangle$. All the knowledge bases contain propositional formulae and are not closed under deduction, and all are stratified by degree of belief as discussed above.

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 [27] and are based on the rules suggested by [20] which, in turn, were based on those in the dialogue game DC introduced by MacKenzie [19]. The rationality rules specify the preconditions for making the move. 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 $CS_i(C) = CS_{i-1}(C)$

⁵ The names stem from the study of persuasion dialogues— P argues “pro” some proposition, and C argues “con”.

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}_{DC}^d . 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 [23], 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.

Clearly this set of moves/locutions defines the communication language CL , and hopefully it is reasonably clear from the description so far how argumentation between agents takes place; a prototypical dialogue might be 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' . and asserts $B_p(\neg q)$.
5. ...

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 [27].

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 proposition ϕ for which there is an acceptable argument (S, ϕ) .

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 [27].

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 proposition ϕ for which there is an argument (S, ϕ) if no stronger rebutting argument exists.
- a *skeptical* agent can accept any proposition 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 [25, 26]. 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 the that agent is aware, 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', ψ) 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', ψ) , 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 [25, 26]. 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 a plan which clashes in some way (we could easily extend the protocol so that B does not have to put forward this plan, but can instead engage A in a persuasion dialogue about A ’s plan in a way that was not considered in [25, 26]). 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 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 [13].

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

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

Proposition 13. *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 .*

Note that since we can’t determine exactly what H says, and therefore what are the contents of $CS(H)$, 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 [25, 26] 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.

5 Summary

Argumentation-based approaches to inter-agent communication are becoming more widespread, 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 chapter 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 [25].

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.

References

1. L. Amgoud and C. Cayrol. On the acceptability of arguments in preference-based argumentation framework. In *Proc. 14th Conf. Uncertainty in AI*, pages 1–7, 1998.
2. L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and AI*, 34:197–215, 2002.
3. L. Amgoud, N. Maudet, and S. Parsons. Modelling dialogues using argumentation. In E. Durfee, editor, *Proc. 4th Intern. Conf. on Multi-Agent Systems*, pages 31–38, Boston, MA, USA, 2000. IEEE Press.
4. L. Amgoud, N. Maudet, and S. Parsons. An argumentation-based semantics for agent communication languages. In *Proc. 15th European Conf. on AI*, 2002.
5. L. Amgoud and S. Parsons. Agent dialogues with conflicting preferences. In J.-J. Meyer and M. Tambe, editors, *Proc. 8th Intern. Workshop on Agent Theories, Architectures and Languages*, pages 1–15, 2001.
6. L. Amgoud, S. Parsons, and N. Maudet. Arguments, dialogue, and negotiation. In W. Horn, editor, *Proc. 14th European Conf. on AI*, pages 338–342, Berlin, Germany, 2000. IOS Press.
7. L. Amgoud, S. Parsons, and L. Perrussel. An argumentation framework based on contextual preferences. In J. Cunningham, editor, *Proc. Intern. Conf. Pure and Applied Practical Reasoning*, London, UK, 2000.
8. 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.
9. F. Dignum, B. Dunin-Kęplicz, and R. Verbrugge. Creating collective intention through dialogue. *Logic Journal of the IGPL*, 9(2):305–319, 2001.
10. 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.
11. T. Finin, Y. Labrou, and J. Mayfield. KQML as an agent communication language. In J. Bradshaw, editor, *Software Agents*. MIT Press, Cambridge, MA, 1995.
12. FIPA. Communicative Act Library Specification. Technical Report XC00037H, Foundation for Intelligent Physical Agents, 10 August 2001.
13. 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.
14. 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.
15. C. L. Hamblin. *Fallacies*. Methuen, London, UK, 1970.

16. 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, *Proc. 4th Biennial Conf. Ontario Soc. Study of Argumentation (OSSA 2001)*, Windsor, Ontario, Canada, 2001.
17. R. Johnson. *Manifest Rationality: A Pragmatic Theory of Argument*. Lawrence Erlbaum Associates, Mahwah, NJ, USA, 2000.
18. J. A. Levin and J. A. Moore. Dialogue-games: metacommunications structures for natural language interaction. *Cognitive Science*, 1(4):395–420, 1978.
19. J. D. MacKenzie. Question-begging in non-cumulative systems. *J. Philosophical Logic*, 8:117–133, 1979.
20. N. Maudet and F. Evrard. A generic framework for dialogue game implementation. In *Proc. 2nd Workshop on Formal Semantics and Pragmatics of Dialogue*, University of Twente, The Netherlands, May 1998.
21. P. McBurney. *Rational Interaction*. PhD thesis, Department of Computer Science, University of Liverpool, 2002.
22. P. McBurney and S. Parsons. Risk agoras: Dialectical argumentation for scientific reasoning. In C. Boutilier and M. Goldszmidt, editors, *Proc. 16th Conf. on Uncertainty in AI*, Stanford, CA, USA, 2000. UAI.
23. P. McBurney and S. Parsons. Games that agents play: A formal framework for dialogues between autonomous agents. *J. Logic, Language, and Information*, 11(3):315–334, 2002.
24. P. McBurney and S. Parsons. Dialogue game protocols. In Marc-Philippe Huget, editor, *Agent Communications Languages*, Berlin, Germany, 2003. Springer. (This volume.).
25. S. Parsons and N. R. Jennings. Negotiation through argumentation — a preliminary report. In *Proc. 2nd Intern. Conf. on Multi-Agent Systems*, pages 267–274, 1996.
26. S. Parsons, C. Sierra, and N. R. Jennings. Agents that reason and negotiate by arguing. *Logic and Computation*, 8(3):261–292, 1998.
27. S. Parsons, M. Wooldridge, and L. Amgoud. An analysis of formal interagent dialogues. In C. Castelfranchi and W. L. Johnson, editors, *Proc. First Intern. Joint Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS 2002)*, pages 394–401, New York, USA, 2002. ACM Press.
28. C. Reed. Dialogue frames in agent communications. In Y. Demazeau, editor, *Proc. 3rd Intern. Conf. on Multi-Agent Systems*, pages 246–253. IEEE Press, 1998.
29. 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.
30. M. Schroeder, D. A. Plewe, and A. Raab. Ultima ratio: should Hamlet kill Claudius. In *Proc. 2nd Intern. Conf. on Autonomous Agents*, pages 467–468, 1998.
31. 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.
32. M. P. Singh. Agent communication languages: Rethinking the principles. In *IEEE Computer 31*, pages 40–47, 1998.
33. M. P. Singh. A social semantics for agent communication languages. In *Proc. IJCAI’99 Workshop on Agent Communication Languages*, pages 75–88, 1999.
34. K. Sycara. Argumentation: Planning other agents’ plans. In *Proc. 11th Joint Conf. on AI*, pages 517–523, 1989.
35. D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. SUNY Press, Albany, NY, 1995.
36. M. J. Wooldridge. *Reasoning about Rational Agents*. MIT Press, Cambridge, MA, USA, 2000.
37. M. J. Wooldridge. Semantic issues in the verification of agent communication languages. *J. Autonomous Agents and Multi-Agent Systems*, 3(1):9–31, 2000.