Negotiation through argumentation—a preliminary report

Simon Parsons and N. R. Jennings
Department of Electronic Engineering,
Queen Mary and Westfield College,
Mile End Road,
London E1 4NS.

S.Parsons,N.R.Jennings@qmw.ac.uk

Abstract

The need for negotiation in multi-agent systems stems from the requirement for agents to solve the problems posed by their interdependence upon one another. Negotiation provides a solution to these problems by giving the agents the means to resolve their conflicting objectives, correct inconsistencies in their knowledge of other agents' world view, and coordinate a joint approach to domain tasks which benefits all the agents concerned. We propose a framework, based upon a system of argumentation, which permits agents to negotiate to establish acceptable ways to solve problems.

Introduction

An increasing number of software applications are being conceived, designed, and implemented using the notion of autonomous agents. These applications vary from email filtering in which a personalised digital assistant processes incoming mail to highlight important messages and remove irrelevant ones (Maes 1996), through electronic commerce in which agents buy and sell goods on behalf of their user (Chavez & Maes 1996), to large industrial applications in which agents take responsibility for particular parts of the overall process (Jennings *et al.* 1996). In all of these disparate cases, however, the notion of autonomy is used to denote the fact that the software has the ability to decide for itself which goals it should adopt and how these goals should be achieved (Wooldridge & Jennings 1995).

In most agent applications, the autonomous components need to interact with one another because of the inherent interdependencies which exist between them. The predominant mechanism for managing these interdependencies at run-time is negotiation—the process by which a group of agents communicate with one another to try and come to a mutually acceptable agreement on some matter (Bussmann & Müller 1992). Negotiation is so central precisely because the agents are autonomous. For an agent to influence an acquaintance, the acquaintance has to be persuaded that it should act in a particular way. The means of achieving this state are to make proposals, trade options, offer concessions, and (hopefully) come to a mutually acceptable agreement—in other words to negotiate.

This paper presents a well-grounded framework for describing the reasoning process of negotiating agents. This framework is based upon a system of argumentation which may be used both at the level of an agent's internal reasoning and at the level of negotiation between agents. An originating agent puts forward an initial proposal. The recipient agents evaluate the proposal by constructing arguments for and against it. If the proposal is unacceptable, the recipient constructs an argument against the initial proposal or in favour of a new alternative. This process continues until a proposal or counter-proposal is acceptable to all the parties involved or until the negotiation breaks down without an agreement.

This paper presents a formal model covering the essence of the negotiation process which can be specialised to describe specific strategies and tactics, an integrated framework for assessing proposals and for generating appropriate counter-proposals, and an intuitively appealing way of conducting reasoning and negotiation in the presence of imprecise and missing information.

A framework for negotiation

Examination of the literature on negotiation from the fields of social psychology (Pruitt 1981), game theory (Rosenschein & Zlotkin 1994), and distributed AI, (Bussmann & Müller 1992; Lâasri *et al.* 1992) reveals a significant level of agreement on the main stages involved in the process. We use this commonality to underpin our generic negotiation model. In our view, negotiation is a process that takes place between two or more agents who are attempting to achieve goals which they cannot, or prefer not to, achieve on their own. These goals may conflict, in which case the agents have to bargain about which agent achieves which goal, or the agents may depend upon one another to achieve the goals, in which case they only have to discuss how to go about achieving the goals. In either case, the process of negotiation proceeds by the exchange of proposals, critiques, explanations and meta-information.

- A *proposal*, broadly speaking, is some kind of solution to the problem that the agents face. It may be a single complete solution, single partial solution, or a group of complete or partial solutions. A proposal may be made either independently of other agents' proposals, or based on proposals or critiques made by other agents.

  Proposals can be more complex than just suggestions for joint action—they may include suggested trade-offs ("I will help you do this provided that you help me do that") or suggest conditions under which the proposal holds ("I offer you this service under these conditions").

- A *critique* is one of the ways in which an agent responds to a proposal made by another agent. It may just be a remark as to whether or not the proposal is accepted or a comment on which parts of the proposal the agent likes, and which parts it dislikes. A more complex kind of critique is a *counter-proposal*, an alternative proposal which is more favourable to the responding agent than the original.

  The process of generating the critique is the method by which the agent evaluates the proposal, and by returning the critique to the originating agent the responding agent aims to provoke alternative proposals that may be more acceptable. The more information placed in the critique, the easier it is for the original agent to respond appropriately.

- An *explanation* is additional information explaining why a proposal was made that an agent can supply in support of its proposal. For instance, an agent might support its suggestion that another agent help it by pointing out that the goal it is proposing that they both attain will benefit them both (a fact which might not be obvious).

- The role of *meta-information* is to focus the local search by agents for solutions. Thus, by supplying information about why it had a particular objection to a proposal, one agent might help another to focus its search for another, more acceptable, suggestion.

The process of negotiation starts when an agent generates a proposal. Other agents then either accept it, critique it or make counter-proposals. Following this, the original agent then either sends clarifying information that may resolve any problems revealed by the critiques, makes a new proposal, critiques a counter-proposal, or indicates its acceptance of the counter-proposal. This process continues until all the agents involved are, in some sense, happy with a proposal or it is felt that no agreement can be reached. By "happy" it is not meant that this is the optimum proposal from the point of view of the agent, but that it represents an acceptable compromise.

To implement this form of negotiation a number of mechanisms are required by every agent taking part in the negotiation process, over and above those for communication[1]. The main mechanisms required are a means of generating proposals, explanations, critiques and counter-proposals and a means of evaluating proposals. There is also the desirable requirement for a means of generating and evaluating meta-information.

It is our contention that the use of a particular system of argumentation delivers all of these mechanisms, and is therefore a good basis both for building negotiating systems and for providing a theoretically grounded means of describing the process. In the next section we discuss what this system of argumentation is, and how it can be used for negotiation.

Negotiation by dialectic argumentation

The system of argumentation which we use as the basis for negotiation is based upon that proposed by Fox and colleagues (Fox, Krause, & Ambler 1992; Krause *et al.* 1995). As with many systems of argumentation, it works by constructing a series of logical steps (arguments) for and against propositions of interest. It is, however, particularly interesting in that the weight given to an argument may be determined by examining the support for the steps in the argument. Because this closely mirrors the way that human dialectic argumentation (Jowett 1875) proceeds, this system seems a promising basis for building a framework for dialectic argumentation by which agents can negotiate.

A system of argumentation

In classical logic, an argument is a sequence of inferences leading to a conclusion. If the argument is correct, then the conclusion is true. Consider the simple database Δ_1 which expresses some very familiar information in a Prolog-like notation in which variables are capitalised and ground terms and predicate names start with small letters.

$$\begin{array}{ll} f1 : human(socrates). & \Delta_1 \\ r1 : human(X) \rightarrow mortal(X). & \end{array}$$

The argument $\Delta_1 \vdash mortal(socrates)$ may be correctly made from this database because $mortal(socrates)$ follows from Δ_1 given the usual logical axioms and rules of inference. Thus a correct argument simply yields a conclusion which in this case could be paraphrased '$mortal(socrates)$ is true in the context of $f1$ and $r1$'. In the system of argumentation adopted here this traditional form of reasoning is extended to allow arguments to indicate support and doubt in propositions, as well as proving them. The degree of support or doubt is ascertained by examining the propositions used in the derivation, and so those propositions that are used are

[1] For simplicity we assume throughout the paper that agents have a commonly agreed communication protocol, a common means of structuring their messages, and that they are able to reach a common understanding of the domain terms that they exchange.

$$\text{Ax}\ \frac{}{\Delta \vdash_{ACR} (p, \{a\})}\ (p,a) \in \Delta \qquad\qquad \top\text{-I}\ \frac{}{\Delta \vdash_{ACR} (\top, \emptyset)}$$

$$\wedge\text{-I}\ \frac{\begin{array}{c}\Delta \vdash_{ACR} (q, B)\\ \Delta \vdash_{ACR} (p, A)\end{array}}{\Delta \vdash_{ACR} (p \wedge q, A \cup B)} \qquad \vee\text{-E}\ \frac{\begin{array}{c}\Delta \vdash_{ACR} (p \vee q, A)\\ \Delta, (p, A) \vdash_{ACR} (r, B)\\ \Delta, (q, A) \vdash_{ACR} (r, C)\end{array}}{\Delta \vdash_{ACR} (r, B \cup C)} \qquad \neg\text{-E}\ \frac{\begin{array}{c}\Delta \vdash_{ACR} (p, A)\\ \Delta \vdash_{ACR} (\neg p, A)\end{array}}{\Delta \vdash_{ACR} (\bot, A)}$$

$$\wedge\text{-E1}\ \frac{\Delta \vdash_{ACR} (p \wedge q, A)}{\Delta \vdash_{ACR} (p, A)} \qquad \vee\text{-I1}\ \frac{\Delta \vdash_{ACR} (q, A)}{\Delta \vdash_{ACR} (p \vee q, A)} \qquad \neg\text{-I}\ \frac{\Delta, (\neg p, \emptyset) \vdash_{ACR} (\bot, A)}{\Delta \vdash_{ACR} (p, A)}$$

$$\wedge\text{-E2}\ \frac{\Delta \vdash_{ACR} (p \wedge q, A)}{\Delta \vdash_{ACR} (q, A)} \qquad \vee\text{-I2}\ \frac{\Delta \vdash_{ACR} (p, A)}{\Delta \vdash_{ACR} (p \vee q, A)} \qquad \rightarrow\text{-I}\ \frac{\Delta, (p, \emptyset) \vdash_{ACR} (q, A)}{\Delta \vdash_{ACR} (p \rightarrow q, A)}$$

$$\text{EFQ}\ \frac{\Delta \vdash_{ACR} (\bot, A)}{\Delta \vdash_{ACR} (p, A)} \qquad \text{RAA}\ \frac{\Delta, (p, \emptyset) \vdash_{ACR} (\bot, A)}{\Delta \vdash_{ACR} (\neg p, A)} \qquad \rightarrow\text{-E}\ \frac{\begin{array}{c}\Delta \vdash_{ACR} (p, B)\\ \Delta \vdash_{ACR} (p \rightarrow q, A)\end{array}}{\Delta \vdash_{ACR} (q, A \cup B)}$$

Figure 1: Argumentation Consequence Relation

recorded. This form of argumentation may be seen as a formalisation of work on informal logic and argumentation in philosophy (van Eemeren *et al.* 1996), though it should be stressed that it was developed quite independently. It is summarised by the following schema:

$$\text{Database } \vdash_{ACR} (\text{Sentence}, \text{Grounds})$$

where $\vdash_{ACR}$ is a suitable consequence relation. Informally, 'Grounds' is a set of labels[2] denoting the facts and rules used to infer the 'Sentence' which the argument supports. This kind of reasoning is similar to that provided by labelled deductive systems (Gabbay 1992), but it differs in its use of the labels. Whilst most labelled deductive systems use their labels to control inference, this system of argumentation uses the labels to determine which of its conclusions are most valid.

To formalise this kind of reasoning we start with a language. We will take a set of propositions $\mathcal{L}$ which includes both $\bot$, the contradiction, and $\top$, the ever-true proposition. We also have a set of connectives $\{\rightarrow, \neg, \wedge, \vee\}$, and the following set of rules for building the well formed formulae of the language:

- If $l \in \mathcal{L}$ then l is a well formed formula (*wff*).
- If l is a *wff*, then $\neg l$ is a *wff*.
- If l and m are *wff*s then $l \rightarrow m$, $l \wedge m$, and $l \vee m$ are *wff*s.
- Nothing else is a *wff*.

This procedure gives a set of *wff*s, which we will call $\mathcal{W}$. To this language we add a set of modalities, $\mathcal{M} = \{B_i, D_i, It_i\}$ which can be applied to any member of $\mathcal{W}$, and which allow the representation of the beliefs, desires and intentions of agent i[3]. Thus $B_2(x)$ indicates that Agent 2 believes the proposition x. The modalities are exactly those suggested by Kraus *et al.* (Kraus, Nirkhe, & Sycara 1993), and have the same interpretation (which for reasons of space we will not repeat here). The set of all formulae that may be built from $\mathcal{W}$ and the modalities $\mathcal{M}$ we will call $\mathcal{M}(\mathcal{W})$. We then allow agents to have databases built up of pairs $(\langle formula \rangle, \langle label \rangle)$ where $\langle formula \rangle$ is a member of $\mathcal{M}(\mathcal{W})$ and $\langle label \rangle$ is a unique label.

With this formal system, any agent can use the argument consequence relation $\vdash_{ACR}$ given in Figure 1 to build arguments for or against propositions in $\mathcal{L}$ that it is interested in, using the information in its database. For instance, the rule 'Ax' sanctions the deduction of the argument $(p, \{a\})$ from the database fact (p, a). Thus from a labelled fact it is possible to construct an argument for that fact whose grounds are the label. Similarly, the rule '$\rightarrow$-E' sanctions the construction of an argument for q from arguments for p and $p \rightarrow q$, giving q grounds which are the union of those for p and $p \rightarrow q$. This can be considered to be

[2] In practice the grounds of an argument are the actual facts and rules used. However, for the sake of simplicity when writing this paper we often talk of the grounds as just a set of labels which provide a means of identifying the facts and rules used.

[3] These modalities are a common means of describing the behaviour of individual agents which are situated in multi-agent systems. In some cases, these modalities are explicitly represented in the agent architecture (for example in BDI architectures such as dMars (Ingrand, Georgeff, & Rao 1992) and GRATE* (Jennings 1995)). However, it has been argued that all agent-based systems can be abstractly characterised using these modalities even if they are not explicitly present (Dennett 1990; Rao & Georgeff 1995). The work described here is neutral with respect to the level at which the system architecture is prescribed by our model.

an argumentation version of the modus ponens rule in classical logic. It should be noted that Figure 1 says nothing about quantification, so that it is not possible to instantiate rules such as $P(x) \rightarrow Q(x)$ given $P(a)$ in order to learn $Q(a)$. This, however, does not prevent us from handling first-order rules—instead, first-order rules are instantiated with every ground instance before arguments are constructed. In other words, if we have $P(x) \rightarrow Q(x)$ and $P(a)$ and $P(b)$, we replace $P(x) \rightarrow Q(x)$ with $P(a) \rightarrow Q(a)$ and $P(b) \rightarrow Q(b)$ before any arguments are constructed. The system of argumentation formed using $\mathcal{M}(\mathcal{W})$ and $\vdash_{ACR}$ will be referred to as $\mathcal{SA}_{\mathcal{M}(\mathcal{W})}$.

Example 1. To see how arguments are built, consider the following simple database in the language of $\mathcal{SA}_{\mathcal{M}(\mathcal{W})}$ which is part of the knowledge of a home-improvement agent (of the kind that we would like to have at home), and which we will call Agent 1:

$$
\begin{array}{ll}
(It_1(Do(agent_1, hang\_picture)), f1) & \Delta_2 \\
(B_1(Have(agent_1, picture)), f2) & \\
(B_1(Have(agent_1, nail)), f3) & \\
(B_1(Have(agent_1, hammer)), f4) & \\
(B_1(Have(W, hammer)) & \\
\quad \wedge B_1(Have(X, nail)) \wedge B_1(Have(Y, picture)) & \\
\qquad \rightarrow B_1(Can(Z, hang\_picture)), r1) & \\
(B_1(Can(X, Y)) \wedge It_1(Do(X, Y)) & \\
\qquad \rightarrow B_1(Do(X, Y)), r2) &
\end{array}
$$

So 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. From this information, Agent 1 can use $\vdash_{ACR}$ to build the following argument:

$$\Delta_2 \quad \vdash_{ACR} \quad (B_1(Do(agent_1, hang\_picture)), \{f1, f2, f3, f4, r1, r2\})$$

indicating that it has an argument for hanging the picture. □

Typically an agent will be able to build several arguments for a given proposition, some of which will be in favour of the proposition, and some of which will be against the proposition (in which case they are for its negation). In order to establish whether or not the set of arguments as a whole are in favour of the proposition, it is desirable to provide some means of *flattening* the set of arguments into some measure of how favoured the proposition is. One way of doing this is to attach a numerical or symbolic weight to arguments and then have a flattening function that combines these in a suitable way. This approach is useful when reasoning under uncertainty (Parsons 1996a), and so promises to be useful for agents which only have partial information about their world. However, it is also possible to be rather more subtle and use the structure of the arguments themselves to determine how good they are. It is this approach that we consider to be most useful from the point of view of negotiation (and it can easily be combined with a weight-based system (Parsons 1996b) to allow negotiation with uncertain information.)

In general, an argument built using $\mathcal{SA}_{\mathcal{M}(\mathcal{W})}$ is a pair (p, A) where p is some proposition in which we are interested, and A is a set of labels. We can identify two important classes of arguments:

Non-trivial argument: An argument (p, A) is non-trivial if the set of facts labelled by A is consistent.

Tautological argument: An argument (p, A) is tautological if $A = \emptyset$.

and the important idea of defeat between arguments:

Defeat: Let (s, A) and (s', B) be arguments from some database Δ where s and s' are sentences of the form $M^1(p_1) \wedge \ldots \wedge M^n(p_n)$ and $M^{1'}(q_1) \wedge \ldots \wedge M^{m'}(q_m)$, respectively, the M^i and $M^{i'}$ being belief, desire or intention modalities. The argument (s', B) can be defeated in one of two ways. Firstly, (s, A) *rebuts* (s', B) if $p_i \equiv \neg q_j$ for some p_i and q_j. Secondly, (s, A) *undercuts* (s', B) if there is $l \in B$ which labels a sentence $M^{1''}(r_1) \wedge \ldots \wedge M^{l''}(r_l)$ and $p_i \equiv \neg r_k$ for some p_i and r_k.

Notions of defeat in argumentation have been widely studied (for instance by (Loui 1987; Vreeswijk 1989; Pollack 1992; 1994; Dung 1995)). The notion that we use here is broadly in line with the consensus on the issue. It is also the natural extension of the notion of defeat proposed by Elvang-Gøranssen *et al.* (Elvang-Gøransson, Krause, & Fox 1993) to the multi-agent case. The difference is as follows. The notion of defeat proposed by Elvang-Gøranssen *et al.* would recognise the conflict between $B_1(a)$ and $\neg B_1(a)$, but would not identify the conflict between $B_1(a)$ and $B_2(\neg a)$. Our extension, by virtue of the fact that it looks inside the modalities, is able to detect this latter type of defeat.

Now, using this idea of defeat along with those of non-triviality and tautology, we can flatten a set of arguments for and against a proposition. We do this by classifying all the arguments in the set into classes of acceptability. We have, in order of increasing acceptability:

A1 The class of all arguments that may be made from Δ.

A2 The class of all non-trivial arguments that may be made from Δ.

A3 The class of all arguments that may be made from Δ for propositions for which there are no rebutting arguments.

A4 The class of all arguments that may be made from Δ for propositions for which there are no undercutting arguments.

A5 The class of all tautological arguments that may be made from Δ.

Informally, the idea is that arguments in more acceptable classes are less questionable than those in lower classes. Thus, if we have an argument for a proposition a which is in class A2, and an argument against a which is in A1, then the result of flattening the pair of arguments is that there is an overall argument for a. Since any argument from any class is included in all classes of lower acceptability, there is an order over the acceptability classes defined by set inclusion:

$$A_5(\Delta) \subseteq A_4(\Delta) \subseteq A_3(\Delta) \subseteq A_2(\Delta) \subseteq A_1(\Delta)$$

so that arguments in smaller classes are more acceptable than arguments in larger classes since there is less reason for thinking that there is something wrong with them (because, for instance, there is no argument which rebuts them).

Example 2. Considering the example of our house-proud agent once again, we can see that its argument for hanging the picture is in class $A_4(\Delta_2)$ since it is non-trivial, has no rebutting or undercutting arguments made against it, but is not tautological. □

Negotiation as argumentation

The next point to address is how negotiation by argumentation proceeds, considering, for simplicity, just the two-agent case[4]. The first step is the construction of an argument $(It_i(a), \Delta_i)$ by one of the agents, Agent i, for at least one of its intentions $It_i(a)$ using some subset Δ_i of its total knowledge base. Agent i then passes its argument $(It_i(a), \Delta_i)$ to the other agent, Agent j[5]. Having received this argument, Agent j then examines $(It_i(a), \Delta_i)$ to see if it agrees with the suggestion. The simplest case is when Agent j can find no reason to disagree with the suggestion, and so simply responds with a message to indicate its agreement. More interesting cases occur when Agent j does not agree with the suggestion, and there are several situations in which this may happen.

The first situation is that in which the suggestion directly conflicts with j's objectives. This state of affairs is detected when j can build an argument $(It_j(\neg a), \Delta_j)$ where Δ_j is a subset of its knowledge base. In other words this kind of conflict occurs when j can build an argument that rebuts the initial suggestion. If this happens, then it wants exactly the opposite to Agent i and its only real alternative is to make some completely new suggestion. It does this by constructing an argument $(It_j(b), \Delta'_j)$ and sending this to Agent i, along with $(It_j(\neg a), \Delta_j)$, the latter to show why it disagrees with the initial suggestion.

The second kind of conflict is less severe and occurs when Agent j does not reject the suggestion made by i, but one of the steps by which the suggestion is reached (in other words it can build an undercutting argument for a). This may occur because it conflicts with one of Agent j's intentions, or because in constructing the suggestion, Agent i made an incorrect assumption about one of j's beliefs. In this case, the suggestion can be agreed upon so long as Agent i can find an alternative way of achieving a. To inform i of this, j sends back its undercutting argument $(B_j(\neg c), \Delta_j)$ where $c \in \Delta_i$. Agent i can then examine the combined argument and try to find an alternative argument for a. This can be resubmitted to j and re-evaluated. If Agent j can find no inconsistencies between this new suggestion and its intentions, the suggestion is acceptable and negotiation ends.

The third kind of conflict is when all the steps in Δ_i are acceptable (Agent j cannot rebut any of them), but if Agent j agrees to the suggestion, it will be unable to achieve one of its objectives. This might be, for instance, if i's suggestion involves j using some scarce resource which is then unavailable for carrying out one of j's objectives. Thus if we have two home improvement agents, one with a hammer and one with a nail, the first might suggest that they hang a picture using the nail. The second agent might not object to anything in the plan (it has no objection to hanging pictures or using its nail), but the use of its only nail will stop it carrying out its objective of hanging a mirror.

If there are such conflicts, then Agent j sends Agent i its arguments for the objective (assuming that there is just one for simplicity) $(It_j(d), \Delta_j)$ where there is some e such that $e \in \Delta_j$ and $\Delta_i \cup \Delta_j \vdash_{ACR} \neg e$. Agent i can, as ever, respond to this information by making a totally new suggestion, but by making this reply j allows i to make a related new proposal for some new objective that will satisfy them both. Such a proposal might be $(It_i(a) \wedge It_j(d), \Delta_i \cup \Delta_j \cup \Delta_{i'})$ in which both the original objective and the new one mentioned by j can be achieved using some new resource suggested by i. In the case of the two home improvement agents this might be that the mirror could be hung using a screw held by i.

Considering this kind of negotiation process, it is clear that it falls within the framework suggested above. Firstly it provides a means of generating proposals by constructing arguments for an agent's intentions. This construction process also has the effect of generating an explanation—the explanation for why a particular proposal is made is the argument that sup-

[4] It should be stressed that the limitation to the two-agent case is purely pragmatic in that it makes the description easier. There is no obvious reason why the procedure described here cannot be extended to an arbitrarily large number of agents.

[5] Of course, an agent need not pass the grounds for its proposal to other agents if this would not be in its interests, but if it does, negotiation is likely to be completed more quickly.

ports it. Once the proposal is made, it is evaluated by other agents through argumentation by the device of seeing whether the argument for the proposal may be defeated, either because it rebuts an objective, or because it undercuts or is undercut by the argument for achieving the other agent's objective. If this kind of objection is detected, the argument created by the other agent will serve as a critique of the initial proposal. A counter-proposal may then be generated by either the original agent or the responding agent using the information that form the proposing and/or critiquing arguments as a guide to what is likely to be acceptable. Thus the use of argumentation also provides a mechanism for providing meta-information.

Note that, as described so far, negotiation does not appear to require the need to determine the acceptability of arguments, but this is not so. The acceptability classes are necessary for two reasons. Firstly, they are the means that an agent uses to determine how strongly it objects to proposals. If, when evaluating a proposal, the agent discovers the proposal falls into classes A4 or A5 then it is accepted. If the proposal falls into class A3, then it is a suggestion that might well be accommodated since the suggestion is the second or third type discussed above. If the proposal falls into class A1 or A2 then there is something seriously wrong with it, and a completely new proposal is indicated. The second use of acceptability classes is to evaluate proposals internally before sending them as suggestions. Clearly it is sensible for an agent to vet its proposals to ensure that they are not detrimental to it, and the acceptability class mechanism provides a way of rating possible suggestions to ensure that only the best is sent.

Example 3. To show how this procedure might work in practice, consider the extension of our previous example to the case in which there are two home-improvement agents with different objectives 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:

$$\begin{array}{ll}
(It_1(Do(agent_1, hang_picture)), f1) & \Delta_3 \\
(B_1(Have(agent_1, picture)), f2) & \\
(B_1(Have(agent_1, screw)), f3) & \\
(B_1(Have(agent_1, hammer)), f4) & \\
(B_1(Have(agent_1, screwdriver)), f5) & \\
(B_1(Have(agent_2, nail)), f6) & \\
(B_1(Have(W, hammer)) & \\
\quad \wedge B_1(Have(X, nail)) \wedge B_1(Have(Y, picture)) & \\
\qquad \to B_1(Can(Z, hang_picture)), r1) & \\
(B_1(Have(W, screwdriver)) & \\
\quad \wedge B_1(Have(X, screw) \wedge B_1(Have(Y, mirror)) & \\
\qquad \to B_1(Can(Z, hang_mirror)), r2) & \\
(B_1(Can(X,Y)) \wedge It_1(Do(X,Y)) & \\
\qquad \to B_1(Do(X,Y)), r3) &
\end{array}$$

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

$$\begin{array}{ll}
(It_2(Do(agent_2, hang_mirror)), f7) & \Delta_4 \\
(B_2(Have(agent_2, mirror)), f8) & \\
(B_2(Have(agent_2, nail)), f9) & \\
(B_2(Have(W, hammer)) & \\
\quad \wedge B_2(Have(X, nail)) \wedge B_2(Have(Y, mirror)) & \\
\qquad \to B_2(Can(Z, hang_mirror)) & \\
\qquad \wedge B_2(\neg Have(X, nail)), r4) & \\
(B_2(Can(X,Y)) \wedge It_2(Do(X,Y)) & \\
\qquad \to B_2(Do(X,Y)), r5) &
\end{array}$$

Agent 1 can work out that it is unable to hang the picture on its own because it is unable to build an argument for $Do(agent_1, hang\_picture)$ without using Agent 2's nail, but it can build an argument for $Do(agent_1, hang\_picture)$ that does include the use of the nail:

$$\Delta_3 \quad \vdash_{ACR} \quad (B_1(Do(agent_1, hang\_picture)), \{f1, f2, f4, f6, r1, r3\})$$

This argument is in A4 since Agent 1 is unable to build any arguments which rebut or undercut it, and it is the proposal that it puts to Agent 2. Agent 2 evaluates this proposal by attempting to critique it by building arguments that conflict with it. It finds that with the additional information that Agent 2 passes about its resources, it can build an argument for hanging its mirror using Agent 1's hammer:

$$\Delta_4 \cup \{f1, f2, f4, f6, r1, r3\} \vdash_{ACR} (B_2(Do(agent_2, hang\_mirror)), \{f4, f7, f8, f9, r4, r5\})$$

and it detects that this argument conflicts with the original proposal since the same supporting information allows it to deduce:

$$\Delta_4 \cup \{f1, f2, f4, f6, r1, r3\} \vdash_{ACR} (B_2(\neg(Have(agent_2, nail)), \{f4, f7, f8, f9, r4, r5\})$$

This second argument undercuts the initial proposal by rebutting $f6$, and so when both arguments are taken together, both are in A3 since both are undercut but neither have their conclusion rebutted. Agent 2 then passes this information back to Agent 1 as a critique of the original proposal. 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:

$$\begin{array}{l}
\Delta_3 \cup \{f7, f8, f9, r4, r5\} \vdash_{ACR} \\
\quad (B_1(Do(agent_1, hang_picture) \\
\quad \wedge B_1(Do(agent_2, hang_mirror)), \\
\quad \{f1, f2, f3, f4, f5, f6, f7, f8, r1, r2, r3, r5\}
\end{array}$$

As far as Agent 1 knows, this argument is in A4 since it may not be rebutted by any argument using information which it knows that Agent 2 has. It therefore passes this argument to Agent 2 as a counter-proposal. When it is critiqued by Agent 2, it indeed turns out to be the case that Agent 2 cannot find any defeating argument for this new proposal and so accepts it. □

Related work

Being such a multi-disciplinary subject, there is a vast literature on the topic of negotiation. Hence our purpose in this section is not to provide a comprehensive overview, but rather to highlight some of the work which is most related to our own.

Bussmann and Müller (1992) draw upon social psychology to devise a negotiation model and algorithm that can be employed by agents in a cooperative environment. Their model is much richer than those found in traditional multi-agent systems (see (Lâasri *et al.* 1992) for a review) and accords well with our generic model. However it lacks a rigorous theoretical underpinning and it assumes that agents are inherently cooperative. Lâasri *et al.* (1992) present a similarly rich negotiation model, although drawn from a predominantly multi-agent systems background, but again make the limiting assumption of cooperating agents.

Rosenschein and Zlotkin's research (Rosenschein & Zlotkin 1994) is representative of a growing body of work on negotiation which is based on game theory. This work does not make the cooperating agent assumption; indeed agents are regarded as self-interested utility maximisers. Despite producing some important results, including some related to deceit and lying in negotiation, their work embodies a number of limiting assumptions. The main concerns are that the agents are assumed to have complete knowledge of the payoff matrix, and hence of the other agents' preferences, and also that precise utility values can be provided. Our approach inherently assumes a partial information perspective and is more qualitative in nature.

Sycara's work on the Persuader system (Sycara 1989) employs argumentation as part of a system that operates in the domain of labour negotiations. Although demonstrating the power and elegance of the approach, her system has a centralised arbitrator to handle the disagreements and is thus less general than ours.

Sycara's work led to subsequent research by Kraus *et al.* (Kraus, Nirkhe, & Sycara 1993) into providing a logical model of the process of argumentation. Their approach involves defining a new logic to define the agent's properties and then identifying five different types of argument that can be used in conjunction with their model (threats, rewards, appeals to precedent, appeals to prevailing practice, and appeals to self-interest). Our approach differs in that we adopt a system of argumentation as our start point and put in place the basic infrastructure for using argumentation as the negotiation metaphor. Their five types of argument, and many others besides, could be implemented in our system simply by instantiating different behavioural rules within the individual agents.

Discussion

This paper has introduced a system of argumentation and shown how it can be used to implement a form of dialectic negotiation. There are a number of points that should be made about our system.

The first point is that it is not simply a specific scheme for performing negotiation, but also a general framework which is capable of capturing a broad range of styles of negotiation. Thus it can encompass negotiation with both benevolent and competitive agents. What we have discussed is the mechanism for building and evaluating proposals. The style in which this is done can be altered without changing this general mechanism, just by altering the rules in the knowledge-base (which form the grounds of the arguments) rather than the rules for argument construction (which define the argument consequence relation). As a result, changing from the kind of co-operative agents described here to ones whose proposals contain lies and threats, for instance, can be done without altering the basic negotiation mechanism which will still be based around the construction of arguments and the passing of proposals. In other words, negotiation in different environments can be carried out using different rules from the knowledge-base whilst still employing the same theorem prover, a facility which makes this system of negotiation very flexible.

The second point is that that a theorem prover exists for constructing arguments (Krause, Ambler, & Fox 1992). Thus we already have a means of implementing this system of negotiation. However, a good deal more work needs to be done on the theorem prover to ensure suitable efficiency, by constraining the construction of arguments so that they concern relevant propositions and by providing tractable means of classifying those arguments that are built into acceptability classes.

The third point is that the generality of this system of argumentation suggests that an agent equipped with it will gain more than just a natural mechanism for negotiation. As mentioned above, argumentation has proved a useful mechanism for handling uncertain information, and this kind of ability will be necessary for any agent operating in the real world. Indeed, argumentation can be seen as a form of due-process reasoning of the kind that Hewitt (1985) argued would be required by any truly open system.

The fourth point is that although the example that we have discussed to illustrate the use of argumentation in this paper was fairly simple, that does not mean that the kind of negotiation that our system can perform is particularly limited. It is not. More complex examples can easily be handled by the system, we just feared that they would be rather opaque to the reader. This form of argumentation has proved itself powerful enough to handle a number of real-world applications of reasoning under uncertainty (Fox 1996), and can easily be extended if additional expressiveness is required, by changing the underlying language. Indeed, we are already working on the extension necessary to handle temporal information.

References

Bussmann, S., and Müller, H. J. 1992. A negotiation framework for cooperating agents. In *Proceedings of the CKBS-SIG Conference*.

Chavez, A., and Maes, P. 1996. Kasbah: An agent marketplace for buying and selling goods. In *Proceedings of the 1st International Conference on The Practical Application of Intelligent Agents and Multi-Agent Technology*.

Dennett, D. C. 1990. *The Intentional Stance*. Cambridge, MA.: MIT Press.

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

Elvang-Gøransson, M.; Krause, P.; and Fox, J. 1993. Dialectic reasoning with inconsistent information. In *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*.

Fox, J.; Krause, P.; and Ambler, S. 1992. Arguments, contradictions and practical reasoning. In *Proceedings of the 10th European Conference on Artificial Intelligence*.

Fox, J. 1996. A unified framework for hypothetical and practical reasoning (2): lessons from clinical medicine. In *Proceedings of the International Conference on Formal and Applied Practical Reasoning*.

Gabbay, D. 1992. LDS—labelled deductive systems. 7th Expanded Draft, Imperial College.

Hewitt, C. 1985. The challenge of open systems. *Byte* 10(4):223–242.

Ingrand, F. F.; Georgeff, M. P.; and Rao, A. S. 1992. An architecture for real-time reasoning and system control. *IEEE Expert* 7(6).

Jennings, N. R.; Corera, J.; Laresgoiti, I.; Mamdani, E. H.; Perriolat, F.; Skarek, P.; and Varga, L. Z. 1996. Using ARCHON to develop real-word DAI applications for electricity transportation management and particle accelerator control. *IEEE Expert* (to appear).

Jennings, N. R. 1995. Controlling cooperative problem solving in industrial multi-agent systems using joint intentions. *Artificial Intelligence* 75:195–240.

Jowett, B. 1875. *The Dialogues of Plato*. Oxford: Oxford University Press, 2nd edition.

Kraus, S.; Nirkhe, M.; and Sycara, K. 1993. Reaching agreements through argumentation: a logical model (preliminary report). In *Proceedings of the Workshop on Distributed Artificial Intelligence*.

Krause, P.; Ambler, S.; and Fox, J. 1992. ATP user manual. Technical Report 187, Advanced Computation Laboratory, Imperial Cancer Research Fund.

Krause, P.; Ambler, S.; Elvang-Gøransson, M.; and Fox, J. 1995. A logic of argumentation for reasoning under uncertainty. *Computational Intelligence* 11:113–131.

Lâasri, B.; Lâasri, H.; Lander, S.; and Lesser, V. 1992. A generic model for intelligent negotiating agents. *International Journal of Intelligent and Cooperative Information Systems* 1:291–317.

Loui, R. 1987. Defeat among arguments: a system of defeasible inference. *Computational Intelligence* 3:100–106.

Maes, P. 1996. Agents that reduce work and information overload. *Communications of the ACM* 37(7):31–40.

Parsons, S. 1996a. Defining normative systems for qualitative argumentation. In *Proceedings of the International Conference on Formal and Applied Practical Reasoning*.

Parsons, S. 1996b. Representational problems in modelling genetic map data. In *Proceedings of the International Conference on Information Processing and the Management of Uncertainty*.

Pollack, J. L. 1992. How to reason defeasibly. *Artificial Intelligence* 57:1–42.

Pollack, J. L. 1994. Justification and defeat. *Artificial Intelligence* 67:377–407.

Pruitt, D. G. 1981. *Negotiation Behaviour*. London: Academic Press.

Rao, A., and Georgeff, M. 1995. BDI agents: From theory to practice. In *Proceedings of the 1st International Conference on Multi-Agent Systems*.

Rosenschein, J. S., and Zlotkin, G. 1994. *Rules of Encounter*. Cambridge, MA: MIT Press.

Sycara, K. 1989. Argumentation: Planning other agents' plans. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*.

van Eemeren, F. H.; Grootendorst, R.; Henkemans, F. S.; Blair, J. A.; Johnson, R. H.; Krabbe, E. C. W.; Plantin, C.; Walton, D. N.; Willard, C. A.; Woods, J.; and Zarefsky, D. 1996. *Fundamentals of Argumentation Theory: A Handbook of Historical Backgrounds and Contemporary Developments*. Mahwah, NJ: Lawrence Erlbaum Associates.

Vreeswijk, G. 1989. The feasibility of defeat in defeasible reasoning. In *Proceedings of the 1st International Conference on Knowledge Representation and Reasoning*.

Wooldridge, M. J., and Jennings, N. R. 1995. Intelligent agents: Theory and practice. *The Knowledge Engineering Review* 10:115–152.