Automated Negotiation: Prospects, Methods and Challenges

N. R. Jennings[1], P. Faratin[2], A. R. Lomuscio[3], S. Parsons[4], C. Sierra[5] and M. Wooldridge[4]

[1] Dept. of Electronics and Computer Science, University of Southampton,
Southampton SO17 1BJ, UK.
nrj@ecs.soton.ac.uk

[2] Dept. of Electronic Engineering, Queen Mary and Westfield College,
University of London, London E1 4NS, UK.
p.faratin@qmw.ac.uk

[3] Dept. of Computing, Imperial College of Science, Technology and Medicine,
London SW7 2BZ, UK.
a.lomuscio@doc.ic.ac.uk

[4] Dept. of Computer Science, University of Liverpool,
Liverpool L69 7ZF, UK.
{s.d.parsons, m.j.wooldridge}@csc.liv.ac.uk

[5] Artificial Intelligence Research Institute, Spanish Scientific Research Council,
Campus UAB, 08193 Bellaterra, Barcelona, Spain.
sierra@iiia.csic.es

1. Introduction

An increasing number of computer systems are being viewed in terms of *autonomous agents*. There are two main drivers to this trend. Firstly, agents are being advocated as a next generation model for engineering complex, distributed systems [13], [44]. Secondly, agents are being used as an overarching framework for bringing together the component AI subdisciplines that are necessary to design and build intelligent entities [24], [32]. While there is still much debate about the precise nature of agenthood, an increasing number of researchers find the following characterisation useful [44]:

> *an agent is an encapsulated computer system that is situated in some environment and that is capable of flexible, autonomous action in that environment in order to meet its design objectives*

There are a number of points about this definition that require elaboration. Agents are: (i) clearly identifiable problem solving entities with well-defined boundaries and interfaces; (ii) situated (embedded) in a particular environment—they receive inputs related to the state of their environment through sensors and they act on the environment through effectors; (iii) designed to fulfill a specific purpose—they have particular objectives (goals) to achieve; (iv) autonomous—they have control both over their internal state and over their own behaviour; (v) capable of exhibiting flexible problem solving behaviour in pursuit of their design objectives—they need to be both reactive (able to respond in a timely fashion to changes that occur in their environment) and proactive (able to act in anticipation of future goals) [45].

When adopting an agent-oriented view of computation, it is readily apparent that most problems require or involve *multiple agents*: to represent the decentralised nature of the problem, the multiple loci of control, the multiple perspectives and/or the competing interests [6]. Moreover, these agents will need to interact with one another, either to achieve their individual objectives or to manage the dependencies that follow from being situated in a common environment [7],

[14]. These interactions can vary from simple information interchanges, to requests for particular actions to be performed and on to cooperation (working together to achieve a common objective) and coordination (arranging for related activities to be performed in a coherent manner). However, perhaps the most fundamental and powerful mechanism for managing inter-agent dependencies at run-time is *negotiation*—the process by which a group of agents come to a mutually acceptable agreement on some matter. Negotiation underpins attempts to cooperate and coordinate (both between artificial and human agents) and is required both when the agents are self interested and when they are cooperative. It is so central precisely because the agents are autonomous. For an agent to influence an acquaintance, the acquaintance needs to be convinced 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 short, to negotiate.

Given its ubiquity and importance in many different contexts, negotiation theory incorporates a broad range of phenomena and makes use of many different approaches (e.g. from AI, Social Psychology and Game Theory). Despite this variety, however, automated negotiation research can be considered to deal with three broad topics (see [21] for a more detailed classification scheme):

- *Negotiation Protocols*: the set of rules that govern the interaction. This covers the permissible types of participants (e.g. the negotiators and any relevant third parties), the negotiation states (e.g. accepting bids, negotiation closed), the events that cause negotiation states to change (e.g. no more bidders, bid accepted) and the valid actions of the participants in particular states (e.g. which messages can be sent by whom, to whom, at what stage).

- *Negotiation Objects*: the range of issues over which agreement must be reached. At one extreme, the object may contain a single issue (such as price), while on the other hand it may cover hundreds of issues (related to price, quality, timings, penalties, terms and conditions, etc.). Orthogonal to the agreement structure, and determined by the negotiation protocol, is the issue of the types of operation that can be performed on agreements. In the simplest case, the structure and the contents of the agreement are fixed and participants can either accept or reject it (i.e. a take it or leave it offer). At the next level, participants have the flexibility to change the values of the issues in the negotiation object (i.e. they can make counter-proposals to ensure the agreement better fits their negotiation objectives). Finally, participants might be allowed to dynamically alter (by adding or removing issues) the structure of the negotiation object (e.g. a car salesman may offer one year's free insurance in order to clinch the deal).

- *Agents' Decision Making Models*: the decision making apparatus the participants employ to act in line with the negotiation protocol in order to achieve their objectives. The sophistication of the model, as well as the range of decisions that have to be made, are influenced by the protocol in place, by the nature of the negotiation object, and by the range of operations that can be performed on it.

The relative importance of these three topics varies according to the negotiation and environmental context. Thus, in some circumstances the negotiation protocol is the dominant concern (e.g. [33], [43]). For example, the system designer may determine that the negotiation is best organised using a particular form of auction (e.g. English, Dutch, Vickrey, First-Price Sealed Bid). This mechanism design choice constrains the types of operations that can be performed on the negotiation object (no counter-proposals or issue extensions) and prescribes the behaviour of the agents' decision making models (e.g. strategic behaviour is pointless and the agents' dominant strategy is to simply bid up to their true reservation value). In other cases, however, the agent's decision making model is the dominant concern (e.g. [37], [40]). Here, the protocol does not prescribe an optimal strategy for the agent and there is scope for strategic reasoning to determine the best course of action. In such cases, the relative success of two agents is determined by the effectiveness of their reasoning model—the better the model, the greater the agent's reward.

Given the wide variety of possibilities, it should be clear that there is no universally best approach or technique for automated negotiation. Rather, there is an eclectic bag of methods with properties and performance characteristics that vary significantly depending on the negotiation context. The aim of this paper is, therefore, to examine the space of negotiation opportunities for autonomous agents, to identify and evaluate some of the key techniques, and to highlight some of the major challenges for future automated negotiation research. This paper is *not* meant as a survey of the field of automated negotiation. Rather, the descriptions and assessments of the various approaches are generally undertaken with particular reference to work in which the authors have been involved. However, the specific issues raised should be viewed as being broadly applicable.

The remainder of this paper is structured as follows. Section 2 presents a generic framework for automated negotiation. This framework is then used to structure the subsequent discussion and analysis of the various negotiation techniques; section 3 deals with game theoretic techniques, section 4 with heuristic techniques, and section 5 with argumentation-based techniques. Finally, section 6 outlines some of the major challenges that need to be addressed before automated negotiation becomes pervasive.

2. A Generic Framework for Automated Negotiation

Negotiation can be viewed as a distributed search through a space of potential agreements (figure 1). The dimensionality and topology of this space is determined by the structure of the negotiation object. Indeed, one could consider each attribute of the negotiation object to have a separate dimension associated with it; clearly, in this view, the space of figure 1 concerns two attributes. Thus, when new issues are added (or old ones removed) during the course of a negotiation, then extra dimensions are added (or removed) and the number of points of agreement may increase (or decrease). Similarly, if an agent changes one of the values of one of the attributes within an offer, it is moving from one point in the agreement space to another.

For a given negotiation, the participants are the active components that determine the direction of the search. At the start of this process, each agent has a portion of the space in which it is willing to make agreements. Typically, it also has some means of rating the points in its space and some means of using this rating to determine the actual agreements it makes. Negotiation proceeds by the participants suggesting specific points (or regions) in the agreement space as potentially acceptable. During the negotiation process, the participants' agreement spaces (as well as their rating functions) may change: they may expand, contract, or shift, for example because their environment changes or because they are persuaded to change their views. The search terminates when the required number of participants find a mutually acceptable point in the agreement space or when the protocol dictates that the search should be terminated (for whatever reason) without making an agreement.

Given this metaphor, figure 1 can be seen to represent agent A1 negotiating with two other agents (A2 and A3). The agreement structure is the same in both cases. The current offer in the A1-A3 negotiation (these interactions are the lightly coloured exchanges) is in the area of overlap between the two agents meaning that it could represent a potential agreement between them. However the current offer in the A1-A2 negotiation (these interactions are the darkly coloured exchanges) will not lead to an agreement since it is outside the agreement space of A2 (indeed in this case, A1 and A2 currently have non-intersecting areas of agreement meaning that no deal is possible). For more on this metaphor for viewing the agreement space see [15], [20], [22].

From this representation, it can be seen that the minimal negotiation capabilities are: (i) to propose some part of the agreement space as being acceptable; and (ii) to respond to such a proposal indicating whether it is acceptable. In other words, the minimum requirement of a negotiating agent is the ability to make and respond to proposals. Here we consider a proposal to be a solution to the negotiation problem; either a single complete proposed solution, a single partial solution, or a group of complete or partial solutions. In terms of the agreement space, these different kinds of proposals become a single point, a region of the space, a set of points, or a set of regions of the space (for example a partial solution would be any region of the space in which the quality was above some level and the price below a certain threshold). Generally speaking, proposals can be made either independently of other agents' proposals or based on the negotiation history.

Arguably the simplest kind of negotiation we can imagine is a Dutch auction [31]. The auctioneer (one agent in the negotiation) calls out prices (negotiation objects with a single attribute). When there is no signal of acceptance from the other parties in the auction (other agents in the negotiation) the auctioneer makes a new offer which it believes will be more acceptable (by reducing the price). Here, because of the convention (protocol) under which the auction operates, a lack of response is sufficient feedback for the auctioneer to infer a lack of acceptance. However in anything more complex than this rather special case, the minimal requirement for the "other agents" is that they are able to indicate dissatisfaction with proposals that they find unacceptable.

If agents can only accept or reject others' proposals, then negotiation (and especially negotiation over objects that are multi-dimensional) can be very time consuming and inefficient since the proposer has no means of ascertaining why the proposal is unacceptable, nor whether the agents are close to an agreement, nor in which dimension/direction of the agreement space it should move next. Hence the proposer is essentially picking points in the agreement space based on its perception of what others prefer and hoping that it will eventually stumble upon something acceptable. To improve the efficiency of the negotiation process, the recipient needs to be able to provide more useful feedback on the proposals it receives. This feedback can take the form of a *critique* (comments on which parts of the proposal the agent likes or dislikes[1]) or a *counter-proposal* (an alternative proposal generated in response to a proposal). From such feedback, the proposer should be in a position to generate a proposal that is more likely to lead to an agreement (if it chooses to do so).

Consider the concept of a critique first. A critique provides two forms of feedback: (i) it suggests constraints on particular negotiation issues and (ii) it indicates acceptance/rejection of particular parts of the proposal (or indeed of the whole proposal). To illustrate these points, consider the following short dialogues that are examples of proposals followed by critiques:

```
A: I propose that you provide me with service X under the following
   conditions.
B: I am happy with the price of X, but the delivery date is too late.
```

1. To avoid introducing an unnecessarily large number of different types of statement, we consider simple accept/reject statements to be special cases of critiques.

```
A: I propose that I will provide you with service Y if you provide me with
     service X.
B: I don’t want service Y.
```

In the first case, the critique indicates those aspects of the proposal that are acceptable and those that need to be modified and it also suggests a constraint on one of the issues (delivery date earlier than the current suggestion). In the second case, the critique indicates outright rejection of part of the proposal. Generally speaking, the more information placed in the critique, the easier it is for the original agent to determine the boundaries of its opponent’s agreement space.

Counter-proposals are the second feedback mechanism. A counter-proposal is simply a proposal that is more favourable to the sender, made in response to a previous proposal. The following are examples of proposals followed by counter-proposals:

```
A: I propose that you provide me with service X.
B: I propose that I provide you with service X if you provide me with
     service Z.

A: I propose that I provide you with service Y if you provide me with
     service X.
B: I propose that I provide you with service X if you provide me with
     service Z.
```

In the first case, the counter-proposal extends the initial proposal, and in the second case it amends part of the initial proposal. Counter-proposals differ from critiques in that the feedback is less explicit (the recipient of a counter-proposal has to infer the constraints and preferences from the way the proposal is re-constituted), but generally more detailed (since specific regions of the opponent’s agreement space are identified).

On their own, proposals, critiques, and counter-proposals are bald statements of what agents want. Thus, their scope is confined solely to the structure of the negotiation object. While it is perfectly possible to base negotiations on just these object-level constructs (indeed this is precisely what most current negotiation models do), doing so diminishes some of the potential of negotiation technology. For example, it means that agents cannot:

- *Justify* their negotiation stance;

  An agent might have a compelling reason for adopting a particular negotiation stance. For example, a company may not be legally entitled to sell a particular type of product to a particular type of consumer or a particular item may be out of stock and the next delivery might not be until the following month. In such cases, the ability to provide the justification for its attitude towards a particular issue can allow the opponent to more fully appreciate an agent's constraints and behaviour.

- *Persuade* one another to change their negotiation stance;

  Agents sometimes need to actively change their opponents' agreement space, or its rating over that space, in order for a deal to be possible. In such cases, agents seek to construct arguments that they believe will make their opponent look more favourably upon their proposal. Thus, arguments seek to identify opportunities for such change (e.g. a car salesman throws in a stereo with a car to increase the value of the good), create new opportunities for change (e.g. a car salesman adds a new dimension to the rating function by highlighting the car's novel security features) or modify existing assessment criteria (e.g. car salesman gets the buyer to change its evaluation function by convincing him that security is more important than high speed).

In both cases, negotiators are providing *arguments* to support their stance (hence *argumentation-based negotiation*). Thus, in addition to generating proposals, counter-proposals and critiques, the negotiator is seeking to make the proposal more attractive (acceptable) by providing additional meta-level information in the form of arguments for its position. The nature and types of the arguments can vary enormously (see [16] [19] [40] for more details). However, common categories include: threats (failure to accept this proposal means something negative will happen to you), rewards (acceptance of this proposal means something positive will happen to you), and appeals (you should prefer this option over that alternative for some reason). Whatever its precise form, the role of the supporting argument is either to modify the recipient's region of acceptability or its rating function over this region. In so doing, arguments have the potential to increase the likelihood and/or the speed of agreements being reached; for example, if agents prefer arguments that are more likely to lead to an agreement (which requires some metric on the agreement space) it is possible to prove that argumentation leads to quicker agreement [42][2]. In the former case, by persuading agents to accept deals that they may previously have rejected. In the latter case, by convincing agents to accept their opponent's position on a given issue (and to cease negotiating over it).

3. Game Theoretic Models

Game theory is a branch of economics that studies interactions between self-interested agents. Like decision theory, with which it shares many concepts, game theory has its roots in the work of von Neumann and Morgenstern [23]. As its name suggests, the basic concepts of game theory arose from the study of games such as chess. However, it rapidly became clear that the techniques and results of game theory can equally be applied to *all* interactions that occur between

2. Of course poorly designed argumentation systems also have the potential to increase the length of the negotiation unnecessarily, for example if agents keep repeating the same arguments ad infinitum. However, poor design of the other aspects of the negotiation mechanism can have similarly adverse effects and so potentially overly long negotiations are not something specific to argumentation-based negotiation.

self-interested agents. Game theory is relevant to the study of automated negotiation because the participants in such negotiations can reasonably be assumed to be self interested.

The classic game theoretic question asked of any particular multi-agent encounter is: what is the best—most rational—thing an agent can do? In most multi-agent encounters, the overall outcome will depend critically on the choices made by all agents in the scenario. This implies that in order for an agent to make the choice that optimises its outcome, it must reason *strategically*. That is, it must take into account the decisions that other agents may make, and must assume that they will act so as to optimise their own outcome. In negotiation, this means, for example, taking into account the private valuations that agents have of the negotiation issues, their private deadlines for making a deal, and so on. Game theory gives us a way of formalising and analysing such concerns.

Negotiation and bargaining were studied in the game theory literature well before the emergence of multi-agent systems as a research discipline, and even before the advent of the first digital computer. However, computer science brings two important considerations to the game theoretic study of negotiation and bargaining:

1. Game theoretic studies of rational choice in multi-agent encounters typically assume that agents are allowed to select the best strategy from the space of all possible strategies, by considering all possible interactions. It turns out that the search space of strategies and interactions that needs to be considered has exponential growth, which means that the problem of finding an optimal strategy is in general computationally intractable. In computer science, the study of such problems is the domain of computational complexity theory [26]. There is a significant literature devoted to the development of efficient (polynomial time) algorithms for apparently intractable problems,

and the application of such techniques to the study of multi-agent encounters is a fruitful ongoing area of work.

2. The emergence of the Internet and World-Wide Web has provided an enormous commercial imperative to the further development of computational negotiation and bargaining techniques [25].

Given a particular negotiation scenario that will involve automated agents, game theoretic techniques can be applied to two key problems:

1. The design of an appropriate *protocol* that will govern the interactions between the negotiation participants. The protocol defines the "rules of encounter" between agents [33]. Formally, a protocol is a set of norms that constrain the proposals that the negotiation participants are able to make. It is possible to design protocols so that any particular negotiation history has certain desirable properties—this is *mechanism design*, and is discussed in more detail below.

2. The design of a particular *strategy* (the agents' decision making models) that individual agents can use while negotiating—an agent will aim to use a strategy that maximises its own individual welfare. A key difficulty here is that, typically, the strategies that work best in theory tend to be computationally intractable, and are hence unusable by agents in practice.

As noted above, mechanism design involves the design of protocols for governing multi-agent interactions, such that these protocols have certain desirable properties. Possible properties include, for example [35] p204:

- *Guaranteed success*: A protocol guarantees success if it ensures that, eventually, agreement is certain to be reached.

- *Maximising social welfare*: Intuitively, a protocol maximises social welfare if it ensures that any outcome maximises the sum of the utilities of negotiation participants. If the utility of an outcome for an agent was simply defined in terms of the amount of money that agent received in the outcome, then a protocol that maximised social welfare would maximise the *total* amount of money “paid out”.

- *Pareto efficiency*: A negotiation outcome is said to be Pareto efficient if there is no other outcome that will make at least one agent better off without making at least one other agent worse off. Intuitively, if a negotiation outcome is not Pareto efficient, then there is another outcome that will make at least one agent happier while keeping everyone else at least as happy.

- *Individual rationality*: A protocol is said to be individually rational if following the protocol— “playing by the rules”—is in the best interests of negotiation participants. Individually rational protocols are essential because without them, there is no incentive for agents to engage in negotiations.

- *Stability*: A protocol is *stable* if it provides all agents with an incentive to behave in a particular way. The best-known kind of stability is *Nash equilibrium*: two strategies `s` and `s'` are said to be in Nash equilibrium if under the assumption that one agent is using `s`, the other can do no better than use `s'`, and vice versa.

- *Simplicity*: A “simple” protocol is one that makes the appropriate strategy for a negotiation participant "obvious". That is, a protocol is simple if using it, a participant can easily (tractably) determine the optimal strategy.

- *Distribution*: A protocol should ideally be designed to ensure that there is no single point of failure (such as a single arbitrator), and ideally, so as to minimise communication between agents.

The fact that even quite simple negotiation protocols can be proven to have such desirable properties as these accounts in no small part for the success of game theoretic techniques for negotiation [17]. As an example, consider the monotonic concession protocol with Zeuthen strategy [33] pp40-49. The monotonic concession protocol for two negotiation participants is as follows. Negotiation proceeds in a sequence of rounds, where at every round, each agent puts forward a proposal. If the proposals "overlap", then agreement has been reached. If the proposals do not overlap, then negotiation proceeds to a further round, where the agents either *make a concession* or else put forward the proposal they made on the preceding round. If neither agent makes a concession, then negotiation terminates with a "conflict deal". This simple protocol ensures that negotiation either monotonically proceeds towards a solution, or else terminates. Now, what strategy should an agent use when faced with such a protocol? One possibility is to use the *Zeuthen strategy*. This strategy essentially says that:

- the agent that should concede is the one with the *most to lose* from the negotiation breaking down;

- the concession that should be made is the minimum required to change the balance of risk—so that the other agent is required to concede on subsequent rounds.

Both of these properties can be formalised quite easily [33] p43. It can be proved that using the Zeuthen strategy when playing the monotonic concession protocol means agents will come to reach agreement on a deal that is Pareto optimal. Unfortunately, the Zeuthen strategy is not in equilibrium; there is an incentive to deviate from the strategy at the last negotiation step [33] p48. Moreover, the strategy is computationally complex, requiring an exponential number of calculations of the agent’s utility function at each negotiation round in order to compute the optimal deal. Nevertheless, it is striking that such a simple and intuitive protocol can be proven to have desirable properties.

Despite these very obvious advantages, however, there are a number of problems associated with the use of game theory when applied to automated negotiation:

- Game theory assumes that it is possible to characterise an agent’s preferences with respect to possible outcomes. *Humans*, however, find it extremely hard to consistently define their preferences over outcomes. In general, human preferences cannot be characterised even by a simple ordering over outcomes, let alone by numeric utilities [32] p475-480. In scenarios where preferences are obvious (such as the case of a person buying a particular CD and attempting to minimise costs), game theoretic techniques may work well. With more complex (multi-issue) preferences, it is much harder to use them.

- The theory has failed to generate a *general* model governing rational choice in interdependent situations [46]. Instead, the discipline has produced a number of highly specialised models that are applicable to specific types of interdependent decision making (e.g. the von Neumann-Morgenstern solution to two-person games [23]). As Binmore notes, in non-cooperative (a sub-branch of game theory) theories:

> ...conclusions (of non-cooperative models) only apply to one specific game. If the details of the rules are changed slightly, the conclusions reached need no longer be valid [5] p. 196.

- Game theory models often assume perfect computational rationality meaning that no computation is required to find mutually acceptable solutions within a feasible range of outcomes. Furthermore, this space of possible deals is often assumed to be fully known by the agents, as is the potential outcome values. This assumption is rarely true in most real world cases; agents typically know their own information space, but they do not know that of their opponent. However, even if the joint space is known, knowing that a solution *exists* is entirely different to knowing what the solution actually *is*. Chess is a classic example of this point. The game has a solution—a strategy for white or black which is either a win or a draw, but the search is computationally complex. Therefore, the notion of perfect rationality, although useful in designing, predicting and proving properties of a system, is not altogether useful in practice. Firstly, it cannot actually be attained: physical mechanisms take time to process information and select actions, hence the behaviour of real agents cannot immediately reflect changes in the environment and will generally be sub-optimal [39]. Secondly, it does not provide a means of analysing the internal design of an agent; one rational agent is as good as another.

Despite these problems, game theory is extremely compelling as a tool for automated negotiation. In cases where it is possible to characterise the preferences and possible strategies of negotiation participants, then game theory has much to offer.

3. Heuristic Approaches

The major means of overcoming the aforementioned limitations of game theoretic models is to use heuristic methods. Such methods acknowledge that there is a cost associate with computation and decision making and so seek to search the negotiation space in a non-exhaustive fashion. This has the effect that heuristic methods aim to produce *good*, rather than *optimal* solutions. The methods themselves may either be computational approximations of game theoretic techniques or they may be computational realisations of more informal negotiation models (e.g. [29], [30]). Examples of such models include: [3], [8], [18], [36], [41]. The key advantages of the heuristic approach can be stated as follows:

- the models are based on realistic assumptions; hence they provide a more suitable basis for automation and they can, therefore, be used in a wider variety of application domains;

- the designers of agents, who are not wedded to game theory, can use alternative, and less constrained, models of rationality to develop *different* agent architectures.

The central concern of this line of work is to model the agent's decision making heuristically during the course of the negotiation (generally speaking, the chosen protocol does not prescribe an optimal course of action). To delve deeper into this area we will concentrate on the heuristic model we have devised; we developed a set of *deliberation mechanisms* that work within a fairly free negotiation protocol [3] [8][9][10] [37].

The space of possible agreements is quantitatively represented by contracts having different values for each issue. Each agent then rates these points in the space of possible outcomes according to some preference structure, captured by a utility function. Proposals and counter-proposals are then offers over single points in this space of possible outcomes, and search terminates either when the time to reach an agreement has been exceeded or when a mutually acceptable solution, a point of intersection of the agents' acceptable outcomes sets, has been reached.

An agent architecture that models the decisions involved in the search for mutually acceptable solutions has also been developed [9]. Whereas the protocol normatively describes the orderings of actions, the decision making mechanisms describe the possible set of agent strategies in using the protocol. These strategies are captured by a negotiation architecture that is composed of responsive and deliberative decision mechanisms. Decision making with the former mechanism is based on a linear combination of simple functions called *tactics*, which manipulate the utility of contracts [8]. The latter mechanisms are subdivided into *trade-off* and *issue manipulation* mechanisms [9]. The former generates offers that manipulate the value, rather than the overall utility, of the offer. The rationale for the trade-off mechanism, like persuasive argumentation, is to make proposals that are more attractive to the opponent. This is achieved not by "providing additional meta-level information" (see section 4), but by providing contracts that are "closer" to the opponent's last offer. The issue manipulation mechanism aims to increase the likelihood of an agreement by adding and removing issues into the negotiation set. The issue

[3.] The negotiation protocol does not prescribe an agent's behaviour, but it does constrain its action selection problem solving through the use of normative rules of interaction. These rules temporally order, according to the agent's roles, communication utterances by specifying both who can say what, as well as when. Specifically, the protocol is a repeated, sequential model where offers are iteratively exchanged. Under this protocol, agents are fully committed to their utterances and utterances are private (unlike, say, the first-price open-cry English auction, where all offers are publicly "heard" by the other interaction participants). The protocol is distributed, symmetric, supports bi and/or multi-agent negotiation as well as distributive and integrative negotiation, involving one or many issues respectively. The utterances (proposals or counter-proposals) are based on previous comments made by other agents, and represent a single complete proposal for a solution. There are no critiques and counter-proposals are based on the object of negotiation (which we term a contract).

manipulation mechanism dynamically alters the structure of the negotiation object, helping to escape local minima in the negotiation dynamics. It does this either by increasing the set of possible outcomes (adding), when negotiation is in deadlock, or, alternatively, removing "noisy" issues that are obstructing the negotiation progress. Since agents have to mutually agree on the set of issues involved in negotiation, the issue manipulation dialogue can be interpreted as a mechanism that modifies the dimensionality of the solution space. The other mechanisms, responsive and trade-off, can then search the altered solution space. When taken together, these three mechanisms represent a continuum of possible decision making capabilities: ranging from behaviours that exhibit greater awareness of environmental resources and less to solution quality, to behaviours that attempt to acquire a given solution quality independently of the resource consumption.

Generally speaking, while heuristic methods do indeed circumvent some of the shortcomings of game theoretic models, they also have a number of comparative disadvantages:

- the models often select outcomes (deals) that are sub-optimal; this is because they adopt an approximate notion of rationality and because they do not examine the full space of possible outcomes;

- the models need extensive evaluation, typically through simulations and empirical analysis, since it is usually impossible to predict precisely how the system and the constituent agents will behave in a wide variety of circumstances.

4. Argumentation-Based Approaches

All of the techniques we have discussed so far have been centred on the trading of proposals. Although, as the previous sections demonstrate, this can be done in a very sophisticated way, the various approaches have three main limitations:

- The proposals themselves generally denote single points in the space of negotiation agreements (a single X or O in Figure 1)[4];
- The only feedback that can be made to a proposal is a counter-proposal, which itself is another point in the space, or an acceptance or withdrawal;
- It is hard to change the set of issues under negotiation in the course of a negotiation (which corresponds to changing the negotiation space of Figure 1 by adding new dimensions).

The aim of argumentation-based negotiation is to remove these limitations. The basic idea behind the argumentation-based approach is to allow additional information to be exchanged, over and above proposals. This information can be of a number of different forms, all of which are arguments which explain *explicitly* the opinion of the agent making the argument. Thus, in addition to rejecting a proposal, an agent can offer a critique of the proposal, explaining why it is unacceptable. This has the effect of identifying an entire area of the negotiation space as being not worth exploring by the other agent. Similarly, an agent can accompany a proposal with an argument which says why the other agent should accept it. This latter kind of argument makes it possible to change the other agent's region of acceptability (by altering its preferences), and also provides a means of changing the negotiation space itself—without the ability to argue for the worth of a new element in the negotiation object, the receiving agent would not, in general, have any basis on which to determine its value. This kind of persuasive argumentation does not

4. Though, of course, agents receiving proposals may assume all kinds of *implicit* information on the basis of them.

have to be tied directly to proposals either. For example, in human argumentation, it is possible to make threats or offer rewards, and these kinds of argument can be captured in this approach [19]. As in human argumentation, agents may not be truthful in the arguments that they generate. Thus when evaluating an argument, the recipient needs to assess the argument on its own merits and then modify this by its own perception of the argument's degree of credibility in order to work out how to respond.

Again to provide more details of this broad class of negotiation, we focus on models we have developed. To this end, the way in which argumentation fits into the general negotiation process was defined in [38] where a simple negotiation protocol for trading proposals was augmented with a series of illocutionary moves which allow for the passing of arguments. It is possible to think of the passing of an argument using one of these moves as marking a transition from the negotiation protocol to a separate argumentation protocol which defines the rules of the game for carrying out an argument dialogue (possible protocols for such dialogues have been suggested in [1], [2]). When the argument dialogue terminates, the agents make the reverse transition and pick up the negotiation dialogue once again.

The exact argumentation mechanism we employ is logic-based [27] and builds on work in argumentation as an approach to handling defeasible reasoning. This makes it possible for agents to handling contradictory statements (which frequently occur during arguments) without collapsing into triviality, and allow conflicting arguments to be resolved. Using argumentation in real agents (as opposed to simple collections of logical statements) means handling the complexities of the agents' mental attitudes, communication between agents, and the integration of the argumentation mechanisms into a complex agent architecture.

These issues were discussed in [28], where we showed how to augment a standard model of argumentation to work for agents which reason using beliefs, desires and intentions. We also discussed how to make use of multi-context systems [12], originally proposed as a means of providing efficient theorem provers for modal logics, to integrate argumentation into a belief-desire-intention agent architecture. This latter strand of work was further developed in [34], and this has led to an implementation in which agents negotiate using argumentation in order to construct joint plans.

For the future, two main areas of work remain. The first is in the definition of suitable argumentation protocols, that is, sets of rules that specify how agents generate and respond to arguments based upon what they know. Initial attempts at defining such protocols are given in [1], [2], but as discussed there, it seems that when defining an argumentation protocol, we “hard-wire” in the attitude that a given agent takes when negotiating with others, defining, for instance, when an argument is found to be persuasive, and when its grounds can be questioned. As a result, we may end up with negotiators which are possibly rather inflexible in their argumentation stance (though more flexible than negotiators which cannot argue). Since this seems rather limiting, we need to investigate this area more with the aim of discovering more flexible argumentation protocols than we currently have. The second main area of work is also related to argumentation protocols, and specifically the transition between the underlying negotiation protocol and the argumentation protocol. When is the right time to make this transition, when is it right to start an argument? Clearly it only makes sense to engage in the complex business of argumentation when it will help the negotiation, but we need to translate this high-level notion of “rightness” into some more concrete decision criterion that can be built into our agents.

While these issues still need to be addressed in order to build fully functional agents capable of argumentation-based negotiation, the work described in this section has laid the foundations for building flexible negotiators. Such agents will have the ability to be persuasive and so achieve agreements which non-argumentation based negotiators cannot. However, the problem with such methods is that they add considerable overheads to the negotiation process, not least in the construction and evaluation of arguments. As a result, we imagine that agents which can argue in support of their negotiations will only ever represent a small, though important, class of automated negotiators.

5. Conclusions

This paper has argued that automated negotiation is a central concern for multi-agent systems research. To this end, a generic framework for classifying and viewing automated negotiations has been developed. This framework was then used to discuss and analyse the three main methods of approach that have been adopted to automated negotiation; namely, game theoretic, heuristic and argumentation-based approaches. For each approach, a brief appraisal of its relative merits and drawbacks is presented. This assessment was generally performed in the context of the authors' own models.

It is clear that much research still needs to be performed in the area of automated negotiation. This research obviously includes extending and developing the specific approaches that have been discussed herein and even developing new methods (such as those based on particle dynamics [17], for example). However, there are also a number of broader issues, which, to date, have received comparatively little attention. These include the following. Firstly, the development of a best practice repository for negotiation techniques. That is, a coherent resource that describes which negotiation techniques are best suited to a given type of problem or domain (much like the way that design patterns function in object-oriented analysis and design [11]).

For each entry, the relative strengths and weaknesses need to be enumerated, the underpinning assumptions need to be explicitly stated, and the likely operational characteristics need to be listed. At present, much of this knowledge is implicit and developers tend to simply adopt the technique (or family of techniques) with which they are most familiar. Secondly, work on knowledge elicitation and acquisition for negotiation behaviour needs to be advanced. At present, there is virtually no work on how a user can instruct an agent to negotiate on their behalf. Such instruction needs to convey the broad negotiation attitude that the agent should adopt, the extent to which the agent can negotiate autonomously, and the degrees of freedom that the agent can explore during a negotiation episode. Finally, and related to the previous two points, work on producing predictable negotiation behaviour needs to be developed. This work is needed to ensure that users are comfortable to delegate negotiation decisions to an autonomous piece of software and that when they do so they are sure that the agent will act within its negotiation mandate.

Acknowledgements

This work has been partially supported by the EPSRC under the grant GR/M07076 and by the EU under grant IST-1999-10948.

References

[1] L. Amgoud, N. Maudet, and S. Parsons (2000) "Modelling dialogues using argumentation" *Proceedings of the 4th Conference on Multi-Agent Systems*, Boston, MA, 31-38.

[2] L. Amgoud, S. Parsons, and N. Maudet (2000) "Arguments, dialogue, and negotiation" *Proceedings of the 14th European Conference on Artificial Intelligence*, Berlin, Germany, 338-342.

[3] M. Barbuceanu and W. Lo (2000) “A multi-attribute utility theoretic negotiation architecture for electronic commerce” *Proceedings of 4th Int. Conf. on Autonomous Agents*, Barcelona, Spain, 239-247.

[4] K. Binmore (1990) “*Essays on the foundations of game theory*” Basil Blackwell.

[5] K. Binmore (1992) “*Fun and Games: A Text on Game Theory*” D. C. Heath and Co.

[6] A. H. Bond and L. Gasser (eds.) (1988) *“Readings in Distributed Artificial Intelligence”* Morgan Kaufmann.

[7] C. Castelfranchi (1998) “Modelling social action for AI agents” *Artificial Intelligence* 103 (1-2) 157-182.

[8] P. Faratin, C. Sierra, and N. R. Jennings (1998) “Negotiation Decision Functions for Autonomous Agents” *Int. Journal of Robotics and Autonomous Systems* 24 (3-4) 159-182.

[9] P. Faratin, C. Sierra, N. R. Jennings and P. Buckle (1999) “Designing Responsive and Deliberative Automated Negotiators” *Proc. AAAI Workshop on Negotiation: Settling Conflicts and Identifying Opportunities*, Orlando, FL, 12-18.

[10] P. Faratin, C. Sierra and N. R. Jennings (2000) “Using Similarity Criteria to Make Negotiation Trade-Offs” *Proc. 4th Int. Conf on Multi-Agent Systems*, Boston, USA, 119-126.

[11] E. Gamma, R. Helm, R. Johnson and J. Vlissides (1995) “*Design Patterns: Elements of Reusable Objected-Oriented Software*” Addison Wesley.

[12] F. Giunchiglia and L. Serafini (1994) “Multilanguage hierarchical logics (or: How we can do without modal logics)” *Artificial Intelligence* 65 29-70.

[13] N. R. Jennings (2000) “On Agent-Based Software Engineering” *Artificial Intelligence* 117 (2) 277-296.

[14] N. R. Jennings (1993) “Commitments and Conventions: The Foundation of Coordination in Multi-Agent Systems” *The Knowledge Engineering Review* 8 (3) 223-250.

[15] N. R. Jennings, S. Parsons, C. Sierra and P. Faratin (2000) "Automated Negotiation" *Proc. 5th Int Conf. on Practical Application of Intelligent Agents and Multi-Agent Systems* (*PAAM-2000*), Manchester, UK, 23-30.

[16] M. Karlins and H. I. Abelson (1970) "*Persuasion*" Crosby Lockwood and Son.

[17] S. Kraus (1997) "Negotiation and Cooperation in Multi-Agent Environments" *Artificial Intelligence* 94 79-97.

[18] S. Kraus and D. Lehmann (1995) "Designing and building an automated negotiation agent" *Computational Intelligence* 11 (1) 132-171

[19] S. Kraus, K. Sycara and A. Evenchik (1998) "Reaching agreements through argumentation: a logical model and implementation" *Artificial Intelligence* 104 1-69.

[20] B. Laasri, H. Laasri, S. Lander and V. Lesser (1992) "A generic model for negotiating agents" *Int. Journal of Intelligent and Cooperative Information Systems*, 1(2), 291-317.

[21] A. R. Lomuscio, M. Wooldridge and N. R. Jennings (2000) "A classification scheme for negotiation in electronic commerce" in *Agent-Mediated Electronic Commerce: A European Perspective* (eds. F. Dignum and C. Sierra), Springer Verlag, 19-33.

[22] R. Loui and D. Moore (2000) "Dialogue and Deliberation" *Cognitive Science* (submitted).

[23] J. Von Neumann and O. Morgenstern (1944) "*The Theory of Games and Economic Behaviour*" Princeton University Press.

[24] N. J. Nilsson (1998) "*Artificial Intelligence: A New Synthesis*" Morgan Kaufmann.

[25] P. Noriega and C. Sierra (eds.) (1999) "*Agent Mediated Electronic Commerce*" LNAI Volume 1571, Springer-Verlag.

[26] C. H. Papadimitriou (1994) "*Computational Complexity*" Addison-Wesley.

[27] S. Parsons and N. R. Jennings (1996) "Negotiation Through Argumentation—A Preliminary Report" *Proc. 2nd Int. Conf. on Multi-Agent Systems*, Kyoto, Japan, 267-274.

[28] S. Parsons, C. Sierra and N. R. Jennings (1998) "Agents that reason and negotiate by arguing" *Journal of Logic and Computation* 8 (3) 261-292.

[29] D. G. Pruitt (1981) "*Negotiation Behaviour*" Academic Press.

[30] H. Raiffa (1982) "*The Art and Science of Negotiation*" Harvard University Press.

[31] J. A. Rodriguez-Aguilar, F. J. Martin, P. Noriega, P. Garcia, C. Sierra (1998) "Towards a tesbed for trading agents in electronic auction markets" *AI Communications* 11 5-19.

[32] S. Russell and P. Norvig (1995) "*Artificial intelligence: a modern approach*" Prentice Hall.

[33] J. S. Rosenschein and G. Zlotkin (1994) "*Rules of Encounter*" MIT Press.

[34] J. Sabater, C. Sierra, S. Parsons and N. R. Jennings (1999) "Using multi-context systems to engineer executable agents" *Proc. 6th Int. Workshop on Agent Theories Architectures and Languages*, Orlando, FL. 131-148.

[35] T. W. Sandholm (1999) "Distributed Rational Decision Making" in *Multiagent Systems* (ed. G. Weiss) MIT Press, 201-258.

[36] A. Sathi and M. S. Fox (1989) "Constraint directed negotiation of resource allocation" in L. Gasser and M. Huhns (eds.) *Distributed Artificial Intelligence II*, 163-195, Morgan Kaufmann.

[37] C. Sierra, P. Faratin and N. R. Jennings (1997) "A Service-Oriented Negotiation Model between Autonomous Agents" *Proc. 8th European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, Ronneby, Sweden, 17-35.

[38] C. Sierra, N. R. Jennings, P. Noriega, and S. Parsons (1997) "A Framework for Argumentation-Based Negotiation" *Proc. 4th Int. Workshop on Agent Theories, Architectures and Languages,* Rode Island, USA 177-192.

[39] H. A. Simon (1982) "*Models of Bounded Rationality, Volume 2*" MIT Press.

[40] K. Sycara (1989) “Argumentation: Planning other Agents’ Plans” *Proc 11th Int. Joint. Conf on AI*, Detroit, MI., 517-523.

[41] K. Sycara (1989) “Multi-agent compromise via negotiation” in L. Gasser and M. Huhns (eds.) *Distributed Artificial Intelligence II*, 119-139, Morgan Kaufmann.

[42] F. Tohme (1997) “Negotiation and defeasible reasons for choice”, *Proc AAAI Spring Symposium on Qualitative preferences in deliberation and practical reasoning*, 95-102.

[43] N. Vulkan and N. R. Jennings (2000) “Efficient Mechanisms for the Supply of Services in Multi-Agent Environments” *Int Journal of Decision Support Systems*, 28 (1-2) 5-19.

[44] M. Wooldridge (1997) “Agent-based software engineering” *IEE Proc Software Engineering* 144 (1) 26-37.

[45] M. Wooldridge and N. R. Jennings (1995) “Intelligent agents: theory and practice” The *Knowledge Engineering Review* 10(2) 115-152.

[46] D. Zeng and K. Sycara (1997) “How can an agent learn to negotiate?” in J. Müller, M. Wooldridge and N. R. Jennings, *Intelligent Agents III*, 233-244, Springer Verlag.

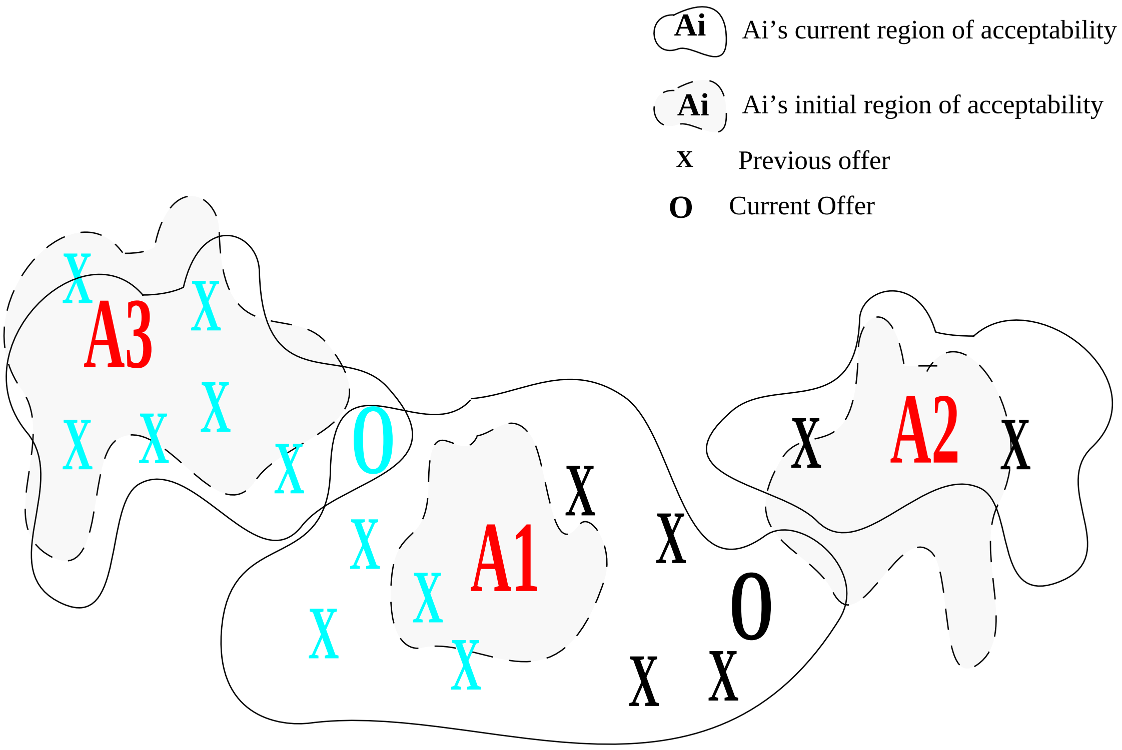

Figure 1: The Space of Negotiation Agreements