The Knowledge Engineering Review, Vol. 18:4, 343–375. © 2004, Cambridge University Press
DOI: 10.1017/S0269888904000098 Printed in the United Kingdom

Argumentation-based negotiation

IYAD RAHWAN[1], SARVAPALI D. RAMCHURN[2], NICHOLAS R. JENNINGS[2], PETER McBURNEY[3], SIMON PARSONS[4] and LIZ SONENBERG[1]

[1]*Department of Information System, University of Melbourne, Parkville 3010, Australia.*
E-mail: i.rahwan@pgrad.unimelb.edu.au, l.sonenberg@dis.unimelb.edu.au
[2]*School of Electronics and Computer Science, University of Southampton, Southampton SO17 1BJ, UK.*
E-mail: sdr01r@ecs.soton.ac.uk, nrj@ecs.soton.ac.uk
[3]*Department of Computer Science, University of Liverpool, Liverpool L69 7ZF, UK.*
E-mail: p.j.mcburney@csc.liv.ac.uk
[4]*Department of Computer and Information Science, Brooklyn College, City University of New York, Brooklyn, NY 11210, USA.*
E-mail: parsons@sci.brooklyn.cuny.edu

Abstract

Negotiation is essential in settings where autonomous agents have conflicting interests and a desire to cooperate. For this reason, mechanisms in which agents exchange potential agreements according to various rules of interaction have become very popular in recent years as evident, for example, in the auction and mechanism design community. However, a growing body of research is now emerging which points out limitations in such mechanisms and advocates the idea that agents can increase the likelihood and quality of an agreement by exchanging *arguments* which influence each others' states. This community further argues that argument exchange is sometimes essential when various assumptions about agent rationality cannot be satisfied. To this end, in this article, we identify the main research motivations and ambitions behind work in the field. We then provide a conceptual framework through which we outline the core elements and features required by agents engaged in argumentation-based negotiation, as well as the environment that hosts these agents. For each of these elements, we survey and evaluate existing proposed techniques in the literature and highlight the major challenges that need to be addressed if argument-based negotiation research is to reach its full potential.

1 Introduction

An increasing number of computer systems are being viewed in terms of multiple interacting autonomous agents. This is because the multi-agent paradigm offers a powerful set of metaphors, concepts and techniques for conceptualising, designing, implementing and verifying complex distributed systems (Jennings, 2001). As a result, applications of agent technology have ranged from electronic trading and distributed business process management to air-traffic and spacecraft control (Parunak, 1999; Wooldridge, 2002).

Here, an agent is viewed as an encapsulated computer system that is situated in an environment and is capable of flexible, autonomous action in order to meet its design objectives (Wooldridge, 1997; Jennings, 2000). In almost all cases, such agents need to interact in order to fulfil their objectives or improve their performance. Generally speaking, different types of interaction mechanisms suit different types of environments and applications. Thus, agents might need mechanisms that facilitate information exchange (Luo *et al*., 2002; Boer *et al*., 2003), coordination (Moulin & Chaib-Draa, 1996; Durfee, 1999) (in which agents arrange their individual activities in

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

a coherent manner), collaboration (Panzarasa *et al.*, 2002; Pynadath & Ttambe, 2002) (in which agents work together to achieve a common objective), and so on. One such interaction that is gaining increasing prominence in the agent community is *negotiation*.

In an attempt to reconcile the definitions proposed by Jennings *et al.*, 2001) and Walton and Krabbe (1995), we offer the following view[1]:

> Negotiation is a form of interaction in which a group of agents, with conflicting interests and a desire to cooperate, try to come to a mutually acceptable agreement on the division of scarce resources.

Automated negotiation among autonomous agents is needed when agents have conflicting objectives and a desire to cooperate. This typically occurs when agents have competing claims on scarce resources, not all of which can be simultaneously satisfied. The use of the word ''resources'' here is to be taken in the broadest possible sense. Thus, resources can be commodities, services, time, money, etc. In short, anything that is needed to achieve something.

In the multi-agent literature, various interaction and decision mechanisms for automated negotiation have been proposed and studied. These include: game-theoretic analysis (Rosenschein & Zlotkin, 1994; Kraus, 2001; Sandholm, 2002); heuristic-based approaches (Faratin, 2000; Kowalczyk & Bui, 2001; Fatima *et al.*, 2002); and argumentation-based approaches (Kraus *et al.*, 1998; Parsons *et al.*, 1998; Sierra *et al.*, 1998). In this paper, we are concerned with argumentation-based approaches. The main distinguishing feature of such approaches is that they allow for more sophisticated forms of interaction than their game-theoretic and heuristic counterparts. This raises a number of research challenges related to both the design of the interaction *environment* as well as the *agents* participating in that interaction.

In this paper, we aim at setting up a research agenda for argumentation-based negotiation in multi-agent systems. We do so by achieving the following. First, we identify the main features of argumentation-based negotiation approaches. We do this by discussing the characteristics of traditional approaches and demonstrate why they fail in particular circumstances due to their underlying assumptions. Second, we discuss, in detail, the essential elements of argumentation-based negotiation frameworks and the agents that operate within these frameworks. We do this by constructing a conceptual model of argumentation-based negotiation, involving external elements (namely, the communication and domain languages, the negotiation protocol, and the information stores) and agent-internal elements (namely, the ability to evaluate, generate, and select proposals and arguments). In the course of discussing each element, we present an overview of existing work in the literature and identify the major challenges and opportunities that remain unaddressed.

This paper is organised as follows. In the next section, we briefly review the different approaches to automated negotiation and outline the contexts in which we believe argumentation-based approaches would be most useful. In Section 3, we describe, in detail, the elements of an argumentation-based framework that are external to the agents, namely the communication and domain languages, the negotiation protocol, and various information stores. In Section 4, we move to discussing the various internal elements and functionalities necessary to enable an agent to conduct argumentation-based negotiation. More precisely, we discuss the processes of argument and proposal evaluation, argument and proposal generation, and argument selection. In Section 5, we summarise the landscape of existing frameworks. Finally, in Section 6, we state the conclusions and summarise the major research challenges.

[1] Note that the precise definition of negotiation is not always stated explicitly in the literature. However, we believe that this definition is a reasonable generalisation of both the explicit and implicit definitions that can be found.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

2 Approaches to automated negotiation

In this section, we discuss the three major classes of approaches to automated negotiation in the multi-agent literature. Even though there may be many ways to classify existing approaches to automated negotiation, the following classification suits our purpose[2].

2.1 Game-theoretic approaches to negotiation

Game theory (Osborne & Rubinstein, 1994) is a branch of economics that studies the strategic interactions between self-interested economic agents[3]. It has its roots in the work of Neuman and Morgenstern (1944). Recently, it has been used extensively to study the interaction between self-interested computational agents (Rosenschein & Zlotkin, 1994; Sandholm, 2002b).

In game-theoretic analysis, researchers usually attempt to determine the optimal strategy by analysing the interaction as a game between identical participants, and seeking its equilibrium (Harsanyi, 1956; Rosenschein & Zlotkin, 1994; Stengel, 2002). The strategy determined by these methods can sometimes be made to be optimal for a participant, given the game rules, the assumed payoffs, and the goals of the participants, and assuming that the participants have no other knowledge of one another than that provided by introspection. Assuming further that participants behave according to the assumptions of rational-choice theory (Coleman, 1990), this approach can guide the design of the interaction mechanism itself, and thus force such agents to behave in certain ways (Varian, 1995; Conitzer & Sandholm, 2002).

However, classical game-theoretic approaches have some significant limitations from the computational perspective (Dash *et al*., 2003). Specifically, most of these approaches assume that agents have unbounded computational resources and that the space of outcomes is completely known. In most realistic environments, however, these assumptions fail due to the limited processing and communication capabilities of the information systems. Agents may be resource-constrained, altruistic, malicious, or simply badly-coded, so that participant behaviour may not conform to the assumptions of rational choice theory[4].

2.2 Heuristic-based approaches to negotiation

To address some of the aforementioned limitations of game-theoretic approaches, a number of heuristic approaches have emerged. Heuristics are rules of thumb that produce good enough (rather than optimal) outcomes and are often produced in contexts with more relaxed assumptions about agents' rationality and resources. The support for particular heuristics is usually based on empirical testing and evaluation (e.g. Faratin, 2000; Kraus, 2001). In general, these methods offer approximations to the decisions made according to game-theoretic studies. One example of this approach is presented by Faratin, Sierra and Jennings in a number of papers (see Sierra *et al*., 1997; Faratin, 2000). In this model, various heuristic decision functions are used for evaluating and generating offers or proposals (i.e., potential deals) in multi-attribute negotiation (Faratin *et al*., 1998). A method for generating tradeoffs is also presented, which aids the construction of alternative offers

[2] For a more comprehensive comparison between the various approaches to automated negotiation, see Jennings *et al*. (2001).

[3] We say ''economic'' agents because economics is concerned with the interaction among people, organisations, etc., rather than among computational agents.

[4] A growing research area in economics that addresses some of the limitations of conventional models is *evolutionary game theory* (Samuelson, 1998), in which the assumption of unbounded rationality is relaxed. In evolutionary models, games are played repeatedly, and strategies are tested through a trial-and-error learning process in which players gradually discover that some strategies work better than others. However, other assumptions, such as the availability of a preference valuation function, still hold. Another attempt is the modelling of ''bounded rationality'' by explicitly capturing elements of the process of choice, such as limited memory, limited knowledge, approximate preferences (that ignore minor differences between options), etc. (Rubinstein, 1997).

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

during bargaining (Faratin *et al.*, 2002). Kowalczyk and Bui (2001) presented a negotiation model with decision procedures based on distributed constraint satisfaction (Yokoo, 1998). This was later extended to allow for multiple concurrent negotiations (Rahwan *et al.*, 2002) and to accommodate fuzzy (as opposed to ''crisp'') constraints (Kowalczyk, 2000). The idea of using fuzzy constraint satisfaction was further investigated by Luo *et al.* (2003). Fatima *et al.* (2001; 2002, 2004) study the influence of information and time constraints on the negotiation equilibrium in a particular heuristic model.

While heuristic methods do indeed overcome some of the shortcomings of game-theoretic approaches, they also have a number of disadvantages (Jennings *et al.*, 2001). Firstly, the models often lead to outcomes that are sub-optimal because they adopt an approximate notion of rationality and because they do not examine the full space of possible outcomes. Secondly, it is very difficult to predict precisely how the system and the constituent agents will behave. Consequently, the models need extensive evaluation through simulations and empirical analysis.

2.3 Argumentation-based approaches to negotiation

Although game theoretic and heuristic based approaches have produced sophisticated systems and are highly suitable for a wide range of applications, they share some further limitations in addition to those mentioned above.

In most game-theoretic and heuristic models, agents exchange *proposals* (i.e. potential agreements or potential deals). This, for example, can be a promise to purchase a good at a specified price in an English auction, a value assignment to multiple attributes in a multi-dimensional auction (Wurman, 1999), or an alternate offer in a bargaining encounter (Larson & Sandholm, 2002). Agents are not allowed to exchange any additional information other than what is expressed in the proposal itself. This can be problematic, for example, in situations where agents have limited information about the environment, or where their rational choices depend on those of other agents[5].

Another limitation of conventional approaches to automated negotiation is that agent's utilities or preferences are usually assumed to be completely characterised prior to the interaction. Thus an agent is assumed to have a mechanism by which it can assess and compare any two proposals. This may be easy, for example, when the utility of the negotiation object is defined in terms of a monetary value, such as the charging rate of a phone call. An agent can compare the proposals of two phone service providers by simply comparing how much they charge per minute. However, in more complex negotiation situations, such as trade union negotiations, agents may well have incomplete information which limits this capability. Thus, agents might:

- lack some of the information relevant to making a comparison between two potential outcomes;
- have limited resources preventing them from acquiring such information;
- have the information, but lack the time needed to process it in order to make the comparison;
- have inconsistent or uncertain beliefs about the environment;
- have unformed or undetermined preferences (e.g., about products new to them); or
- have incoherent preferences.

Thus, to overcome these limitations, the process of acquiring information, resolving uncertainties, revising preferences, etc. often takes place as part of the negotiation process itself.

A further drawback of traditional models to automated negotiation is that agents' preferences over proposals are often assumed to be *proper* in the sense that they reflect the true benefit the agent receives from satisfying these preferences. For example, an agent attempting to purchase a car might assign a high value to a particular brand according to its belief that this brand makes safer

[5] This is typically the case, for instance, with network goods such as fax machines or computer operating systems. Here, the value of a fax machine to one agent depends on whether or not other agents have fax machines.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

cars than other brands. If this belief is false, then the preferences do not properly reflect the agent's actual gain if it was to purchase that car.

Finally, game-theoretic and heuristic approaches assume that agents' utilities or preferences are *fixed*. One agent cannot directly influence another agent's preference model, or any of its internal *mental attitudes* (e.g., beliefs, desires, goals, etc.) that generate its preference model. A rational agent would only modify its preferences upon receipt of new information. Traditional automated negotiation mechanisms do not facilitate the exchange of this information.

Against this background, argumentation-based approaches to negotiation attempt to overcome the above limitations by allowing agents to exchange additional information, or to ''argue''[6] about their beliefs and other mental attitudes during the negotiation process. In the context of negotiation, we view an *argument* as a piece of information that may allow an agent to: (a) *justify* its negotiation stance; or (b) *influence* another agent's negotiation stance (Jennings *et al*., 1998).

Thus, in addition to accepting or rejecting a proposal, an agent can offer a critique of it. This can help make negotiations more efficient. By understanding why its counterpart cannot accept a particular deal, an agent may be in a better position to make an alternative offer that has a higher chance of being acceptable. In a trade union dispute, for example, an agent representing the worker's union might refuse an offer for a modified pension plan made by the organisation's management agent (Sycara, 1985, 1992). As a response, the management agent might offer a different pension plan. If the union agent had been able to explain that the problem with the initial offer was not with its pension plan but rather that it did not include reduced working hours, the management agent would not have bothered exploring different pension plans. Instead, the management agent would have concentrated on finding an arrangement for workload reduction.

Another type of information that can be exchanged is a justification of a proposal, stating why an agent made such a proposal or why the counterpart should accept it. This may make it possible to change the other agent's region of acceptability[7] (Jennings *et al*., 1998), or the nature of the negotiation space itself. For example, an employee negotiating a salary raise might propose a big increase that gets rejected by the manager. After the employee justifies the proposal by denoting her significant achievements during the year, the manager might accept. Agents may also exchange information that results in changing the negotiation object itself, by introducing new attributes (or dimensions) to the negotiation object. For example, the manager might modify the negotiation object such that the negotiation involves not only the salary amount, but also the number of working hours. In this way, the manager might be able to offer reduced working hours instead of a salary increase.

An agent might also make a threat or promise a reward in order to exert some pressure on its counterpart to accept a proposal. For example, a manager requesting a project to be completed by a short deadline might promise a salary raise (or threaten to fire the employees) in order to entice them to allocate more time to working on that particular project[8].

2.4 Summary

From the discussion above it should be clear that there is no universal approach to automated negotiation that suits every problem domain. Rather, there is a set of approaches, each based on

[6] In this survey, we do not treat the topic of argumentation based on defeasible or non-monotonic reasoning as discussed, for example, by Prakken and Vreeswijk (2002), Vreeswijk (1997), Chesnevar *et al.* (2000), Dung (1995) and Loui (1987). Our focus here is on the general characteristics of argumentation in negotiation models for multi-agent systems. One may use either of the above argumentation systems as a basis for an argument-based negotiation system.

[7] The region of acceptability may be defined as the complete set of outcomes the agent is willing to accept.

[8] Promises and threats are also captured in evolutionary game-theoretic models (Samuelson, 1998). For example, by punishing non-cooperative moves by its opponent, an agent sends an indirect threat for future iterations of the game. However, such threats and rewards span over multiple, complete iterations of the same encounter, rather than being part of a single encounter.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

different assumptions about the environment and the agents involved in the interaction. The particular class of approaches that we focus on in this paper, often referred to as *argumentation-based negotiation* (ABN) frameworks, is gaining increasing popularity for its potential ability to overcome the limitations of more conventional approaches to automated negotiation. However, such models are typically more complex than their game-theoretic and heuristic counterparts.

Against this background, the aim of this analytical survey is to identify the main components of an abstract framework for ABN and discuss the different attempts to realise these components. While doing so, we highlight the major challenges encountered in the field.

3 External elements of ABN frameworks

At present, there is no agreed approach to characterising all negotiation frameworks. However, we believe it is instructive to develop such a framework so that the essential components that are needed to conduct automated negotiation, and consequently their associated challenges, can be clearly identified. In this section, we outline those elements that we consider are essential in the design of an ABN framework in particular. By developing an understanding of what an ABN framework is expected to contain, we are in a better position to understand and analyse existing models that have been proposed in the literature. Moreover, this nomenclature enables us to identify the ABN landscape and the main open research questions in the field. In the course of the discussion, we outline some of the major characteristics that differentiate ABN frameworks from other non-argumentation-based approaches to automated negotiation.

Abstractly, a negotiation framework can be viewed in terms of its negotiating agents (with their internal motivations, decision mechanisms, knowledge bases, etc.) and the environment in which these agents interact (with its rules of interaction, communication language, and information stores)[9]. In the remainder of this section, we discuss the main elements that define an ABN framework. In particular, we focus on the elements external to the agent (i.e. those elements that define the environment in which the ABN agents operate and interact). We leave the discussion of the internal features of ABN agents to Section 4.

3.1 Communication language and domain language

Negotiation is, by definition, a form of interaction between agents. Therefore, a negotiation framework requires a language that facilitates such communication (Labrou *et al.*, 1999). Elements of the *communication language* are usually referred to as *locutions*, *utterances* or *speech acts* (Searle, 1969; Traum, 1999). Traditional automated negotiation mechanisms normally include the basic locutions such as `propose` for making proposals, `accept` for accepting proposals, and `reject` for rejecting proposals.

In addition to the communication language, agents often need a common *domain language* for referring to concepts of the environment, the different agents, time, proposals, and so on[10]. When a statement in the domain language is exchanged between agents, it is given particular meaning by the communication language utterance that encapsulates it. For example, in the framework presented by Sierra *et al.* (1998), the locution `offer`$(a, b, \mathit{Price}=\$200 \wedge \mathit{Item}=\mathit{palm130}, t_1)$, means that agent a proposes to agent b, at time t_1 the sale of item *palm130* for the price of \$200. On the

[9] There are other ways in which a negotiation framework can be viewed abstractly, such as those presented by Bartolini *et al.* (2002) and Wurman *et al.* (2001), which view auction frameworks in terms of the rules that parametrise them. However, since these frameworks focus on auction mechanisms, they mainly address the external rules of interaction, and do not address issues such as commitments and preference modification. We believe our model is more suitable for the task at hand because it marks out the features that are peculiar to argumentation-based approaches.

[10] Note that this language may be different from the language used internally by an agent. In such cases, the agent needs to perform some type of *translation* into the common language in order for communication to work (Sierra *et al.*, 1998).

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

Table 1 Differences between ABN and non-ABN frameworks with respect to domain and communication languages

	Non-ABN frameworks	ABN frameworks
Domain language	Expresses proposals only (e.g., by describing products available for sale)	Expresses proposals as well as meta-information about the world, agent's beliefs, preferences, goals, etc.
Communication language	Locutions allow agents to pass call for bids, proposals, acceptance and rejection, etc.	In addition, locutions allow agents to pass meta-information either separately or in conjunction with other locutions

other hand, the `reject` locution gives the same content a different meaning. The locution `reject`(b, a, *Price=$200* $\wedge$ *Item=palm130*, t_2) means that agent b rejects such a proposal made by agent a.

In ABN frameworks, agents need richer communication and domain languages to be able to exchange meta-level information (i.e. information other than that describing outcomes). Therefore, a major distinguishing factor of ABN frameworks is in the type of information that can be expressed and exchanged between agents and, consequently, in the specifications of the agents that generate and evaluate this information. Table 1 shows the main distinguishing features between ABN and non-ABN frameworks as they relate to the communication and domain languages.

3.1.1 State of the art

In existing ABN frameworks, various domain and communication languages have been proposed. They range from those designed as simplistic domain specific languages to more complex languages grounded in rich logical models of agency.

In multi-agent systems, two major proposals for agent communication languages have been advanced, namely the Knowledge Query and Manipulation Language (KQML) (Mayfield *et al.*, 1996) and the Foundation for Intelligent Physical Agents' Agent Communication Language (FIPA ACL) (FIPA, 2001). FIPA ACL, for example, offers 22 locutions. The contents of the messages can be in any domain language. The locution `inform`(a, b, φ, *lan*), for example, allows agent a to inform another agent b of statement φ which is in language *lan*. Other locutions exist allowing agents to express proposals for action, acceptance and rejection of proposals, make various queries about time and place, and so on. FIPA ACL has been given semantics in the form of pre- and post-conditions of each locution. This semantics are based on *speech act theory*, due to a philosopher of language John Austin (Austin, 1962) and his student John Searle (Searle, 1969), in which a locution is seen as an action that affects the world in some way.

While FIPA ACL offers the benefits of being a more or less standard agent communication language, it fails to capture all the utterances needed in a negotiation interaction. For example, FIPA ACL does not have locutions expressing the desire to enter or leave a negotiation interaction, to provide an explicit critique to a proposal or to request an argument for a claim. While such locutions may be constructed by injecting particular domain language statements within locutions similar to those of FIPA ACL, the semantics of these statements fall outside the boundaries of the communication language. Consider the following locution from the framework presented by Kraus *et al.* (1998):

$$\texttt{Request}(j, i, Do(i, \alpha), Do(i, \alpha) \rightarrow Do(j, \beta)).$$

In this locution, agent j requests that agent i performs action α and supports that request with an argument stating that if i accepts, j will perform action β in return. For the locution to properly

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

express a promise, action β must actually be desired by agent *i*. If, in contrast, β is undesirable to *i*, the same locution becomes a threat and might deter *i* from executing α. The locution `Request`, however, does not include information that conveys this distinction.

In order to deal with the above problem, ABN framework designers often choose to provide their own negotiation-specific locutions, which hold the appropriate semantics of the message within them. For example, Sierra *et al.* (1998) and Ramchurn *et al.* (2003b) provide explicit locutions for expressing threats and rewards (e.g., `threaten`(*i*, *j*, α, β) and `promise`(*i*, *j*, α, β)).

Having discussed some issues relating to the communication languages in ABN, let us now discuss the domain languages. In negotiation, the domain language must, at least, be capable of expressing the object of negotiation. In Sierra *et al.*'s model, the domain language can express variables representing negotiation issues (or attributes), constants representing values for the negotiation issues (including a special constant ''?'' denoting the absence of value), as well as equality and conjunction. This enables them to express full or partial multiple-attribute proposals. For example, the sentence

$$(Price = £10) \wedge (Quality = high) \wedge (Penalty = ?)$$

expresses a proposal to agree on a high-quality product or service for the price of *£10*, and with a cancellation penalty yet to be agreed upon. There is also a meta-language for explicitly expressing preferences. For example, the statement $Pref(\ulcorner Price = £10 \urcorner, \ulcorner Price = £20 \urcorner)$ expresses the fact that an agent prefers a price of *£10* to *£20*.

In addition, ABN frameworks may need some way to express plans and resources needed for different plans. This is because agents participating in negotiation may be doing so in order to obtain resources needed for executing their plans. This means that an agent may be able to inform another agent of (parts of) its plans in order to justify its request for particular resources. Sadri *et al.* (2002), for example, express plans using the *plan*(.) predicate. The formula

$$plan(\langle hit(nail), hang(picture) \rangle, \{picture, nail, hammer\})$$

denotes a plan (or intention) to hit a nail and hang a picture. The resources this plan requires are a picture, a nail and a hammer.

Some ABN frameworks also explicitly express information about agents' mental attitudes. The ABN frameworks presented by Kraus *et al.* (1998) and by Parsons *et al.* (1998), for example, allow an agent to represent beliefs about other agents' beliefs, desires, intentions, capabilities, and so on, and are based on logics of belief, desire and intention (BDI) (Rao & Georgeff, 1995; Wooldridge, 2000). An agent can use this information not only in its internal reasoning processes, but also in its interaction with other agents.

The usefulness of the domain language in the context of ABN becomes particularly apparent when agents provide arguments for requesting certain resources, for rejecting certain requests, and so on. The richer the domain language, the richer the arguments that can be exchanged between agents. This will become more evident when we discuss argument generation and evaluation in the following sections.

3.1.2 Challenges

There are a number of challenges in the design of domain and communication languages for ABN. First, there is a need to provide rich communication languages with clear semantics. To this end, Mcburney *et al.* (2003) specified a set of locutions as part of a dialogue game[11] for purchase

[11] Dialogue games are interactions between two or more players, where each player makes a *move* by making some utterance in a common communication language and according to some pre-defined rules. Dialogue games have their roots in the philosophy of argumentation (Aristotle, 1928; Hamblin, 1970). In multi-agent systems, dialogue games have been used to specify dialogue protocols for persuasion (Amgoud *et al.*, 2000a), negotiation (Amgoud & Parsons, 2001), and team formation (Dignum *et al.*, 2000).

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

negotiation among multiple agents. The authors provided public axiomatic semantics to their locutions by stating each locution's externally observable preconditions, the possible response, and the updates to the information and commitment stores[12]. Moreover, the framework presents operational semantics of the whole framework, connecting locutions with each other via the agents' decision mechanisms. However, this framework does not cover the whole spectrum of ABN situations. For example, there are no locutions for explicitly requesting, providing and challenging arguments, or for supporting argumentation over preference criteria. Locutions facilitating argument exchange have been proposed in other frameworks (e.g., Sadri *et al*., 2001a, 2002; Torroni & Toni, 2001; Amgoud *et al*., 2000; Amgoud & Parsons, 2001). There are opportunities for extending the model of Mcburney *et al*. (2003) with a richer argumentation system.

Another prospect of future research is the building of common, standardised domain languages that agent designers can use in order to plug their agents into heterogeneous environments. Efforts towards semantic and syntactic interoperability in domain languages and ontologies, such as the DARPA Agent Markup Language (Hendler & McGuinness, 2000; McGuinness, 2001) and the W3C Web Ontology Language (OWL) (McGuinness & van Harmelen, 2003) are particularly relevant. There is a need for exploring the suitability of these domain languages for supporting ABN and understanding how arguments can be expressed and exchanged.

3.2 Negotiation protocol

Given a communication and domain language, a negotiation framework should also specify a *negotiation protocol* in order to constrain the use of the language. Here we view a protocol as a formal set of conventions governing the interaction among participants (Jennings *et al*., 2001). This includes the *interaction protocol* as well as other rules of the dialogue.

The interaction protocol specifies, at each stage of the negotiation process, who is allowed to say what. For example, after one agent makes a proposal, the other agent may be able to accept it, reject it or criticise it, but might not be allowed to ignore it by making a counterproposal. The protocol might be based solely on the last utterance made, or might depend on a more complex history of messages between agents.

The other rules that form part of the negotiation protocol may address the following issues (Esteva *et al*., 2001; Jennings *et al*., 2001):

- *rules for admission*, which specify when an agent can participate in a negotiation dialogue and under what conditions;
- *rules for participant withdrawal*, which specify when a participant may withdraw from the negotiation;
- *termination rules*, which specify when an encounter must end (e.g., if one agent utters an acceptance locution);
- *rules for proposal validity*, which specify when a proposal is compliant with some conditions (e.g., an agent may not be allowed to make a proposal that has already been rejected);
- *rules for outcome determination*, which specify the outcome of the interaction: in an auction-based framework, this would involve determining the winning bid(s) (Sandholm, 2002a); in argumentation-based frameworks, these rules might enforce some outcome based on the underlying theory of argumentation (e.g., if an agent cannot construct an argument against a request, it accepts it (Parsons *et al*., 1998));
- *commitment rules*, which specify how agents' commitments should be managed, whether and when an agent can withdraw a commitment made previously in the dialogue, how inconsistencies between an utterance and a previous commitment are accounted for, and so on.

In ABN, the negotiation protocol is usually more complex than those in non-ABN. By ''more complex'', we mean that the protocol may involve a larger number of locutions, and a larger

[12] We discuss information and commitment stores in more detail in Section 3.3.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

number of rules. This leads to computational complexity arising from processes such as checking the locutions for conformance with the protocol given the history of locutions.

3.2.1 State of the art

With respect to the interaction protocol, a variety of trends can be found in the ABN literature. Interaction protocols can either be specified in an *explicit* accessible format, or only be *implicit* and hardwired into the agents' specification.

Explicit specification of interaction protocols may be done using finite-state machines (e.g., Sierra *et al*., 1998; Parsons *et al*., 1998). While this approach may be useful when the interaction involves a limited number of permitted locutions, it becomes harder to specify and understand when the number of locutions and their interactions increases significantly. This is particularly problematic when different agent designers need to look up the interaction protocol specification to guide their agents' design and implementation. In such cases, other forms of protocol specification may be more suitable.

Another way of expressing interaction protocols explicitly is using dialogue games (as in, e.g., Amgoud *et al*. (2000), Amgoud and Parsons (2001), McBurney *et al*. (2003)). As mentioned above, dialogue games have the advantage of providing clear and precise semantics of the dialogues, by stating the pre-and post-conditions of each locution as well as its effects on agents' commitments. The following is the specification of a locution from the protocol presented by McBurney *et al*. (2003). This locution allows a seller (or advisor) agent to announce that it (or another seller) is willing to sell a particular option[13].

Locution: `willing_to_sell`(P_1, T, P_2, V), where P_1 is either an advisor or a seller, T is the set of participants, P_2 is a seller and V is a set of sales options.

Preconditions: some participant P_3 must have previously uttered a locution `seek_info` (P_3, S, p) where $P_1 \in S$ (the set of sellers), and the options in V satisfy constraint p.

Meaning: the speaker P_1 indicates to audience T that agent P_2 is willing to supply the finite set $V=\{\overline{a}, \overline{b}, \ldots\}$ of purchase options to any buyer in set T. Each of these options satisfy constraint p uttered as part of the prior `seek`(.) locution.

Response: none required.

Information store updates: for each $\overline{a} \in V$, the 3-tuple (T, P_2, $\overline{a}$) is inserted into $IS(P_1)$, the information store for agent P_1.

Commitment store updates: no effects.

One advantage of dialogue game protocols is that they have public axiomatic semantics. This is because they refer only to observable pre-conditions and effects, rather than to the agents' internal mental attitudes. This makes it easier to verify whether agents are conforming to the protocol.

Other frameworks implicitly hardwire the interaction protocol into the agents' internal specification (e.g., Kraus *et al*., 1998; Sadri *et al*., 2001a,b, 2002; Torroni & Toni, 2001). In these frameworks, the interaction protocol is specified using logical constraints expressed in the form of if–then rules. Since these frameworks describe a logic-based approach to agent specification (Kraus *et al*., 1998) implement their agents using logic programs, while Sadri *et al*. (2001b) use abductive logic programs), the protocol rules are coded as part of the agent's program. These rules take the form $P(t) \wedge C(t) \Rightarrow P'(t+1)$, meaning that if the agent received performative (i.e. locution) P at time t, and condition C was satisfied at that time, then the agent *must* use the performative P' at the next time point. The condition C describes the rationality precondition in the agent's mental state. For example, one rule might state that if an agent received a performative which includes a request for a resource and it does not have that resource, then it must refuse the request. Note that this constitutes private semantics of the protocol, and is hence harder to enforce by an external regulator.

[13] We leave the discussion of ''information stores'' and ''commitment stores'' to Section 3.3.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

The termination rules in negotiation protocols specified as finite-state machines are defined as a set of links to a final state. This is usually the case when one agent utters a `withdrawaccept`(.) locution. In the framework of McBurney *et al*. (2003), a rule specifies that the dialogue ends after an agent utters the locution `withdraw_dialogue`(.) causing either no remaining sellers or no remaining buyers in the dialogue. In some frameworks, however, no termination rules have been defined, and hence the dialogue remains open even after agreement or failure.

In relation to outcome determination rules, some frameworks determine outcomes based on the logical structure of interacting arguments. For example, in the frameworks of Parsons *et al*. (1998) and Amgoud *et al*. (2000), a rule specifies that an agent must accept a request if it fails to produce an argument against that request. A similar case occurs when agents argue about their beliefs – an agent must accept a proposition if it fails to provide an argument for the negation of the proposition. In this sense, outcome determination is *implicit* in the underlying argumentation logic. In other frameworks, such as those of Kraus *et al*. (1998) and Ramchurn *et al*. (2003), outcomes are reached through uttering a specific locution *explicitly* (e.g., by uttering `accept`(.)). Agents may utter such a locution based on some internal utility evaluation.

We leave the discussion of commitment rules to Section 3.3, where we discuss commitment stores.

3.2.2 Challenges

Protocols for ABN share the challenges faced in the design of argumentation protocols in general. For example, there is a need for qualities such as fairness, clarity of the underlying argumentation theory, discouragement of disruption by participants, rule consistency, and so on[14].

One particularly important property is that of *termination*. To this end, some rules for preventing certain causes of infinite dialogues have been proposed. For example, the protocol of Amgoud and Parsons (2001) does not allow agents to repeat the exact same locutions over and over again. The intuition is that this would prevent the agent from, say, repeating the same question over and over again. In subsequent papers, the authors present further analysis of the outcomes of various argumentation-based dialogues (Parsons *et al*., 2002, 2003).

Torroni (2002) studied termination and success in the ABN framework presented earlier (Sadri *et al*., 2001b). Since the ABN framework is grounded in an operationally defined agent architecture based on abductive logic programming, it has been possible to study some properties by referring to the machinery of abduction. In particular, the author determined an upper limit to the maximum length of a dialogue, measured in the number of exchanged messages. Since these results are strongly dependant on the underlying logical system, it is not clear whether these results can be generalised to a variety of protocols without regard to the internal agent architecture.

Another important desired property in ABN protocols is that of *guaranteed success*. Wooldridge and Parsons (2000) investigated the conditions under which particular logic-based negotiation protocols terminate with agreement. They provided results showing the complexity of solving this problem with negotiation frameworks using different domain languages. Most interestingly, they showed that the problem of determining whether a given protocol can be guaranteed to succeed, when used with a FIPA-like communication language, is provably intractable.

An important problem related to interaction protocols in general is that of *conformance checking*. This problem is concerned with answering the question of whether a particular utterance is acceptable, given the history and context of interaction. Conformance checking is one of the sources of complexity in dialogue systems; however, to date, it has received little attention in the ABN literature. Recently, Huget and Wooldridge (2003) investigated applying model checking techniques to this problem.

[14] For a more elaborate discussion of the properties desired in argumentation protocols, refer to (McBurney *et al*., 2002).

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

Another avenue of future research is in the design of admission rules in negotiation protocols. While some frameworks (e.g., McBurney *et al.*, 2003) require that agents explicitly request to enter a negotiation dialogue, to our knowledge, no ABN framework includes external rules that govern admission to the negotiation dialogue. One may envisage situations where only agents with particular credentials, such as reputation or performance history, may be admitted to a negotiation. More work needs to be done on investigating the effect of different admission rules on the outcome of negotiation. For example, a malicious agent may attempt to disrupt the interaction among other participants, and hence should not be admitted. Relevant work has been done in the context of agent admission to electronic institutions (Rodriguez-Aguílar & Sierra, 2002).

3.3 Information stores

In some ABN frameworks, there is no explicit centralised information store available. Instead, agents internally keep track of past utterances (e.g., Kraus *et al.*, 1998). However, in many negotiation frameworks there is a need to keep externally accessible information during interaction. For example, we might need to store the history of utterances for future reference or to store information about the reputation of participants (Rubiera *et al.*, 2001; Yu & Singh, 2002). Moreover, having external information stores makes it possible to perform some kind of *enforcement* of protocol-related behaviours. For example, we may be able to prevent an agent from denying a promise it has previously made.

3.3.1 State of the art

One type of information store that is common in the argumentation literature is the *commitment store*[15]. Commitment stores were initially conceived by Hamblin (1970) as a way of tracking the claims made by participants in dialogue games. Hamblin studied dialogues over beliefs, although he was at pains to state that commitments made in dialogue games should not be construed as necessarily representing the real beliefs of the respective participants (Hamblin, 1970, p. 257). Hamblin's notion of commitment store has been influential in later work on dialogue games, both in philosophy and in multi-agent systems, although the notions of commitment used sometimes differ. In the work on the philosophy of dialogue (e.g., Walton & Krabbe, 1995) the focus is on action commitments, i.e. promises to initiate, execute or maintain an action or course of actions. Commitments to defend a claim if questioned, called propositional commitments, are viewed as special cases of such action commitments by these authors. In the multi-agent systems literature the concern is usually with action commitments, where the actions concerned are assumed to take place outside the agent dialogue. For example, one agent may commit to providing a specified product or service to another agent.

Note that commitment stores should not be confused with the interaction history, which only records the sequence of utterances during the whole interaction[16]. While the latter only form a passive storage of ''unprocessed'' utterances, commitments in commitment stores have more elaborate consequences. For example, when an agent asserts a proposition p, it may not only be committed to believe that p holds, but also to defending that p (if challenged), not denying that p, giving evidence that p, and so on (Walton & Krabbe, 1995). In the multi-agent systems literature, Singh (2000) gave ''social'' semantics for commitments using modal operators in branching-time logic. These semantics are public (i.e. based on external observations of utterances as opposed to agents' internal mental states) and hence can be used for specifying, and checking for conformance with, the interaction protocols. Amgoud *et al.* (2002) also present social semantics of communication based on argumentation. Another difference of commitment stores in comparison with interaction histories is that commitment stores have specific *commitment rules* governing the

[15] For a more detailed discussion of commitments in multi-agent dialogues, see Maudet and Draa (2002).
[16] Sierra *et al.* (1998) use the term *negotiation thread*, while Sadri *et al.* (2001b) use the term *dialogue store*.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

addition and removal of statements that the agent is committed to. One rule may specify, for example, that if the agent retracted a previously asserted claim, it must also retract every claim based on the former via logical deduction. Another relevant concept is that of *pre-commitment* proposed by Colombetti (2000). A request pre-commits the utterer in the sense that the utterer will be committed in case the hearer accepts the request. Commitment stores enable us to capture such pre-commitments.

In the ABN literature, Amgoud and Parsons (2001) define for each agent a publicly accessible commitment store. Adding statements to the commitment store is governed by the dialogue-game rules. For example, when an agent accepts a request for action p, then p is added to its commitment store. Agents may also be allowed to retract commitments under certain conditions. In the context of purchase negotiations, McBurney *et al.* (2003) dealt with the issue of retraction differently. For example, the framework involves two locutions, `agree_to_buy`(.) and `agree_to_sell`(.), for committing to certain resource exchanges. Instead of providing explicit locutions for retracting these commitments, the authors provide additional locutions, `willing_to_buy`(.) and `willing_to_sell`(.), which are softened versions of the former locutions, however, with no commitments incurred (i.e., they are free to refuse to sell or buy something they have previously agreed upon). This way, agents may usefully provide information without necessarily committing to it or having to explicitly retract it.

3.3.2 Challenges

The representation and manipulation of information stores is not a trivial task, and has significant effects on both the performance and outcomes of negotiation dialogues. In particular, information store manipulation rules have a direct effect on the types of utterances agents can make given their previous utterances (i.e., the protocol), the properties of the dialogues (e.g., termination), and the final outcome (e.g., the ability to change one's mind coherently).

Some of the key questions that need to be addressed in an ABN framework are as follows. Under what conditions should an agent be allowed to retract its commitments and how would this affect the properties of dialogues? Under what conditions should an agent be forced to retract its commitments to maintain consistency? While these questions are being investigated in the multi-agent dialogue literature in general (Maudet & Chaib-draa, 2003), there are issues specific to negotiation dialogues. In particular, commitments to providing, requesting, and exchanging resources may require different treatments from commitments in other types of dialogues, such as persuasion or information seeking. Very little work on this problem has been done in existing ABN frameworks.

4 Elements of ABN agents

In the previous section, we discussed the different elements of an ABN framework that are external to the participating agents. Issues such as the interaction protocol, commitment rules, and communication languages represent the environment in which agents operate, but often these say little about how agents are specified, or how they reason about the interaction.

Before we get into a discussion of the general features of an ABN agent, we describe what constitutes (at an abstract level) a basic, non-ABN negotiating agent. This will allow us to clearly contrast the ABN agent from other negotiators, making our analysis more focused. Therefore, we begin by presenting a conceptual model of a simple negotiator in Figure 1. This captures, on a very abstract level, the main components needed by an agent in order to be capable of engaging in negotiation[17]. This model is not meant to be an idealisation of all existing models in the literature, but rather a useful starting point for illustrating how ABN agents differ from other types of agent.

[17] For a more detailed discussion of the conceptual architectures for negotiating agents, refer to (Ashri *et al.*, 2003).

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

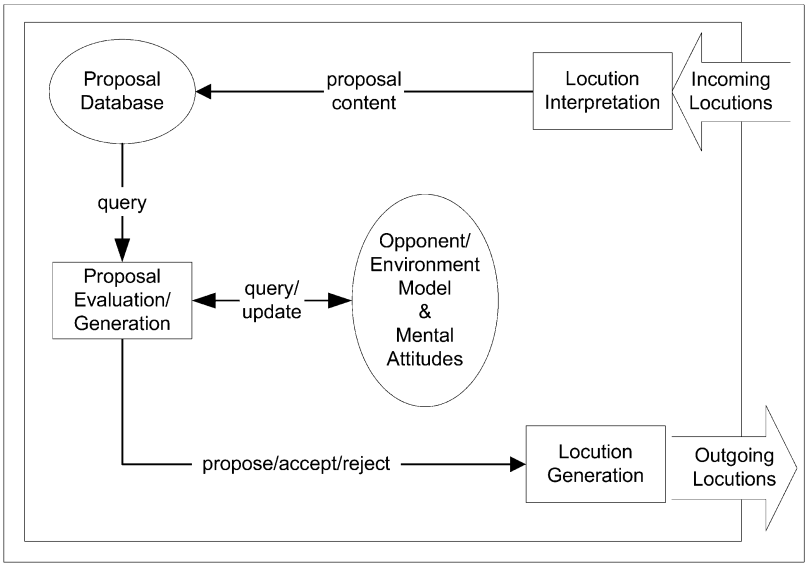


Figure 1 Conceptual elements of a classical negotiating agent

We refer to an agent involved in negotiation interactions which largely depend on exchanging proposals, such as auction-based and bargaining agents, as a classical negotiating agent. This agent needs to have a *locution interpretation* component, which parses incoming messages. These locutions usually contain a proposal, or an acceptance or rejection message of a previous proposal. They might also contain other information about the interaction, such as the identity of the sender (especially in the case of multi-party encounters). Acceptance messages usually terminate the encounter with a deal. A proposal may be stored in a *proposal database* for future reference. Proposals (or rejections) feed into a *proposal evaluation and generation component*, which ultimately makes a decision about whether to accept, reject or generate a counterproposal, or even terminate the negotiation[18]. This finally feeds into the *locution generation* component which sends the response to the relevant party or parties.

A more sophisticated classical agent may maintain a *knowledge base* of its mental attitudes (such as beliefs, desires, preferences, and so on (Wooldridge, 2002)), as well as models of the environment and the negotiation counterpart(s). This knowledge may be used in the evaluation and generation of proposals by judging the validity and worth of the proposals made (for example, by verifying whether proposals are actually feasible and do not conflict with the current observations of the environment). Moreover, the knowledge base may be updated in the light of new information. However, the updates that can be made are somewhat limited because the only information usually available to the agent during the interaction is:

1. proposals (or bids) from the counterpart or a competitor;
2. a message rejecting a proposal initially made by the agent;
3. other observations from the environment (e.g., a manufacturing plant agent bidding for raw material may monitor customer demand changes and bid accordingly).

The agent may be able to infer certain things from this information. For example, by receiving a rejection the agent may infer that the counterpart does not rate certain attribute/value assignments highly. Similarly, by receiving a proposal (or by observing the proposal of another competing

[18] Note that the way components operate is constrained by the negotiation protocol. For example, in English auctions, which are highly *asymmetric*, one agent usually only receives bids while others only generate them. In bargaining models (e.g., Kowalczyk & Bui, 2001), however, the protocol tends to be *symmetric*, allowing both agents to evaluate and generate proposals.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

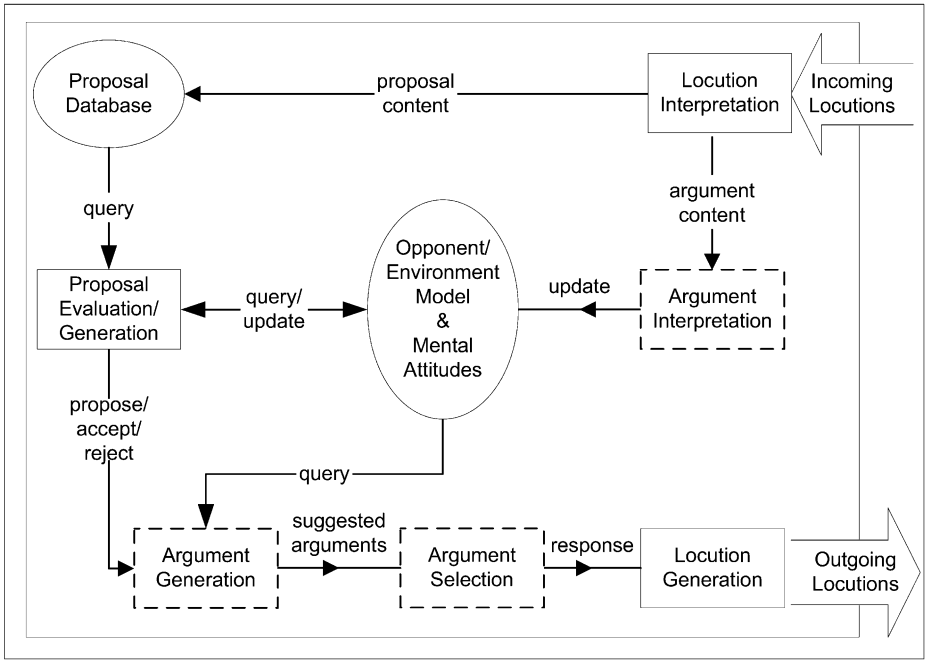


Figure 2 Conceptual elements of an ABN agent (the dashed lined boxes represent the additional components necessary for ABN agents)

bidder) the agent might infer attribute values that appeal to the counterpart (or competitor), which can then guide his own bargaining or bidding strategy[19].

In contrast with a classical negotiating agent, more sophisticated meta-level information can be explicitly exchanged between the ABN agents (see Figure 2)[20]. This, in turn, can have a direct effect on the agent's knowledge base. Therefore, in addition to evaluating and generating proposals, an agent capable of participating in ABN must be equipped with mechanisms for *evaluating* arguments (and updating the mental state accordingly) and for *generating* and *selecting* arguments.

If the locution contains an argument, an *argument evaluation* or *interpretation* mechanism is invoked which updates the agent's mental state accordingly. This may involve updating the agent's mental attitudes about itself and/or about the environment and its counterparts. Now, the agent can enter the *proposal evaluation* stage in the light of this new information. Note that at this stage, not only does the agent evaluate the most recent proposal, but it can also re-evaluate previous proposals made by its counterparts; these proposals are stored in the proposal database. This is important since the agent might (intentionally or otherwise) be persuaded to accept a proposal it has previously rejected.

As a result of evaluating proposals, the agent may generate a counterproposal, a rejection, or an acceptance. In addition, however, a final *argument generation* mechanism is responsible for deciding what response to actually send to the counterpart and what (if any) arguments should accompany the response. For example, the proposal evaluation and generation component might decide that a proposal is not acceptable and the argument generation mechanism might accompany the rejection with a critique describing the reasons behind the rejection. Such arguments might also be explicitly requested by the other party or even enforced by the protocol. Note that an agent may also choose

[19] Similar issues have been investigated in the study of *signalling* in game-theory (Spence, 1974).

[20] Note that the actual way in which ABN agents are designed or implemented may differ from the above. For example, the agent might perform certain operations in a different order, or might combine or further decompose certain functionalities. Therefore, our conceptual model is to be taken in the abstract sense and should not be seen as a prescriptive account of how ABN agents *must* precisely look. Instead, it provides a useful point of departure for beginning an analysis of the generic features of these agents.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

to send a stand-alone argument (i.e. not necessarily in conjunction with a proposal, acceptance or rejection).

At times, there might be a number of potential arguments that the agent can send. For example, in order to exert pressure on a counterpart, an agent might be able to either make a threat or present a logical argument supporting some action. Deciding on which argument to actually send is the responsibility of an *argument selection* mechanism. Finally, this information is given to the *locution generation* mechanism which places this information in the proper message format and utters it.

In summary, negotiating agents must, at least, be able to:

1. interpret incoming locutions;
2. evaluate incoming proposals;
3. generate outgoing proposals;
4. generate outgoing locutions.

An ABN agent needs, *in addition*, to be able to:

1. evaluate incoming arguments and update its mental state accordingly;
2. generate candidate outgoing arguments;
3. select an argument from the set of available arguments.

Now that we have given an overview of the features of an ABN agent, we consider each of these features in more detail. In the course of doing so, we evaluate the state of the art and outline major challenges and opportunities.

4.1 Argument and proposal evaluation

Recall that an ABN agent needs to evaluate potential agreements proposed by its counterpart(s). The agent also needs to be able to evaluate arguments intended at influencing its mental state. While proposals may be evaluated more straightforwardly through comparison with some subjective preference criteria, argument evaluation is less trivial.

Argument evaluation is a central topic in the study of argumentation, and has been studied extensively by philosophers from at least the days of Aristotle (Aristotle, 1928; Hitchcock, 2002). In artificial intelligence, argument evaluation and comparison has been applied, for example, in internal agent deliberation (Kakas & Moraitis, 2003), in legal argumentation (Prakken & Sartor, 2001), and in medical diagnosis (Krause *et al*., 1995; Fox & Parsons, 1998).

Here, however, we find it useful to distinguish between two types of considerations in argument evaluation.

1. *Objective considerations.* An argument may be seen as a *tentative proof* for some conclusion. Hence, an agent, or a set of agents, may evaluate an argument based on some *objective convention* that defines how the quality of that proof is established. This may be done, for example, by investigating the correctness of its inference steps, or by examining the validity of its underlying assumptions. For example, Elvang-Gøransson *et al*. (1993β) proposed a classification of arguments into *acceptability classes* based on the strength of their construction. Arguments may also be evaluated based on their relationships with other arguments. For Dung (1995), for instance, an argument is said to be *acceptable* with respect to a set S of arguments if every argument attacking it is itself attacked by an argument from that set. The set S is said to be *admissible* if it is conflict free and all its arguments are acceptable with respect to S.
2. *Subjective considerations.* Instead of applying an objective, agent-independent convention for evaluating arguments, an agent may choose to consider its own preferences and motivations in making that judgement, or those of the intended audience. In the framework presented by Bench-Capon (2001), for example, different participants in a persuasion dialogue have different preferences over the ''values'' of arguments. Argument assessment and comparison would then take place in accordance with the preferences of the dialogue participants. This means that the

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

participants may influence the outcome of the argument evaluation process by having different *subjective* preferences.

Let us now examine the usage of the above considerations in different types of argumentation dialogues. If two agents are reasoning about what is true in the world (i.e. if they are conducting *theoretical reasoning*), then it makes sense for them to adopt an objective convention that is not influenced by their individual biases and motivations. For example, whether it is sunny outside should not be influenced by whether participants want it to be sunny, but rather only by the material evidence available.

If, on the other hand, two participants are engaged in a dialogue for deciding what course of action to take (i.e. if they are conducting *practical reasoning*), what division of scarce resources to agree upon or what goals to adopt, then it would make more sense for them to consider their subjective, internal motivations and perceptions, as well as the objective truth about their environment[21]. Even objective facts may be perceived differently by different participants, and such differences in perception may play a crucial role in whether or not participants are able to reach agreement. For example, a potential airline traveller may perceive a particular airline as unsafe, while the staff of the airline itself may consider it to be safe. Presumably such a difference in perceptions may be resolved with recourse to objective criteria (if any can be agreed) regarding relative crash statistics, deaths-per-mile-flown on different airlines, etc. However, if, for example, potential travellers perceive a particular airline as unsafe compared with other airlines, despite objective evidence showing the airline to be safer than others, this perception may inhibit them from flying the airline anyway. The marketing team of the airline concerned, in trying to persuade potential travellers to fly with it, will have to engage in dialogue with potential customers on the basis of those customers' subjective perceptions, even though such perception may be false. For the marketers to ignore such mis-perceptions risks the dialogue terminating without the potential customers flying the airline.

In summary, agents participating in negotiation are not concerned with establishing the truth *per se*, but rather with the satisfaction of their needs. Hence, negotiation dialogues require agents to perform argument evaluation based on objective as well as subjective criteria[22]. In other words, agents need to perform *argument evaluation* as part of, or in relation to, *proposal evaluation*.

4.1.1 State of the art

As we argued above, argument evaluation in negotiation must involve both objective as well as subjective considerations, and hence must involve some subjective assessment of proposals put forward by negotiation counterparts. In this subsection, we show some approaches to proposal and argument evaluation in the existing ABN literature.

One approach to proposal and argument evaluation is to assume agents are benevolent, using the following simple normative rule: *if I do not need a resource, I should give it away when asked.* This approach can be found in a number of frameworks (e.g., Parsons *et al.*, 1998; Amgoud *et al.*, 2000b; Sadri *et al.*, 2001b).

Consider the following example from Parsons *et al.* (1998). An agent a intending to hang a picture would produce, after executing its planning procedure, intentions to acquire a nail, a hammer and a picture. Interactions with other agents are only motivated in case the agent is not able to fulfil its intentions on its own. Suppose the agent does not have a nail. This leads the agent to adopt a new intention (we can call that a sub-intention) to acquire a nail, which may be written $I_a(Have(a, nail))$. If a believes that another agent b has a nail, it would generate another sub-sub-intention that b gives the nail to it, written $I_a(Give(b, a, nail))$. This triggers a request to be

[21] Refer to Rahwan *et al.* (2003c) for a related comparison between argumentation over goals and beliefs.

[22] Note that objective argument evaluation may also take into account certain ''preferences'', such as the trust the evaluator has in the agent proposing the argument. However, this remains aimed at establishing the truth, rather than being influenced by the agent's personal gain.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

made to agent b in the form $H_1 \vdash I_a(Give(b, a, nail))$, where the argument H_1 constitutes the sequence of deductive steps taken to reach the request[23]. In general, agent b accepts the request unless it has a conflict with it. There are two types of conflict that would cause b to reject the request as follows.

1. Agent b has a conflicting intention. In argumentation terms, the agent refuses the proposal if it can build an argument that *rebuts* it.
2. Agent b rejects one of the elements of the argument supporting the intention that denotes the request. In argumentation terms, the agent refuses the proposal because it can build an argument that *undercuts* it.

We explain the above two cases using the same picture-hanging example. An example of the first case is if agent b rejects the proposal because it also needs the nail, say to hang a mirror (i.e. it can build an argument for the intention $I_b\neg\ Give(b, a, nail)$)). This argument is based on (among other things) the intention $I_b(Can(b, hang(mirror)))$. An example of the second case is if, in the plan supporting the intention $I_a(Give(b, a, nail))$, agent a made the false assumption that b possesses a nail, written $B_a(Have(b, nail))$. If b actually does not have a nail, then it would adopt the intention of modifying that belief, i.e. $I_b(\neg\ B_a(Have(b, nail)))$. Agents continue through a process of argument exchange, which may involve recursively undercutting each other's arguments until a resolution is reached.

In order for argumentation to work, agents must be able to *compare* arguments. This is needed, for example, in order to be able to reject ''weak'' arguments. Parsons *et al.* compare arguments by classifying them into *acceptability classes* based on the strength of their construction (Elvang-Gøransson *et al.*, 1993b). If two conflicting arguments belong to the same class, the authors assume the agent has some capability to perform comparisons based on utility analysis. However, they do not specify how this decision procedure is actually undertaken, nor do they specify the conditions it needs to satisfy.

A similar approach is taken by Sadri *et al.* (2001b). This framework, however, does not involve arguing about beliefs. If an agent a receives a request for a resource and needs that resource for achieving some goal g_a, the agent rejects the request, unless an alternative acceptable plan for achieving g_a can be produced by the counterpart, with a promise to provide any missing resources for that plan to a. Agents are also assumed to have some ordering over plans that allow them to choose between alternative plans.

In the frameworks of Parsons *et al.* and Sadri *et al.* described above, argument and proposal evaluation take into account a very simplistic subjective rule; that is, to accept any request that the agent does not *currently* need. While this may be useful for facilitating cooperative behaviour and making sure agents preserve their current subjective interests, it may not be suitable in open agent systems where agents may be purely self-interested and may refuse to provide any resources without something in return.

An alternative trend in proposal and argument evaluation is to explicitly take into account the utility of the agent. The basic idea is that the agent would calculate the *expected utility* in the cases where it accepts and rejects a particular proposal. By comparing the expected utilities in these two cases (i.e. in the resulting *states*), the agent would be able to make a decision about whether to accept or reject the proposal. In the framework of Kraus *et al.* (1998), the agent makes a decision about whether to accept a request by evaluating three factors: (i) the Collision_Flag, which fires if the requested action conflicts with one of the agent's intentions; (ii) the Convincing_Factor, which is a value between 0 and 1 assigned to the argument using some *ad hoc* rule (e.g., an appeal to a past promise is assigned a value of 1 if the agent believes it has actually made such promise, and assigned 0 otherwise); and (iii) the Acceptability_Value, which involves a numerical calculation of utility

[23] Note that the argument (or plan) may not contain only intentions, but also belief and desire formulae about the agent and its environment. For example, in the argument H_1, agent a may state the assumption that it believes b has a nail, and so on.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

tradeoffs in the case of accepting the request versus rejecting it. However, it is not clear, from the paper, how these factors are combined to produce a final decision.

Ramchurn *et al.* (2003b) built on the work of Kraus *et al.* (1998) and took it further by factoring the *trust* the agent has in its counterpart when calculating the expected values. The probability of moving into the proposed state is captured by the uncertainty expressed in the trust value. The different factors taken into consideration are combined using fuzzy reasoning mechanisms.

Sierra *et al.* (1998) introduced authority as a criteria for evaluating arguments. They presented an *authority graph* imposed by a relation over agent roles. This graph can be used to specify, for each pair of agents, which agent has higher authority. The authors also propose a way of comparing the authority levels of *sets of* agents. This can be used to compare arguments involving statements made by multiple agents. An argument H_1 is preferred to another H_2 if and only if the authority level of agents involved in H_1 is higher than those in H_2. As an example, the authors define a conciliatory agent, which accepts appeal-to-authority arguments regardless of the content of the justification of the appeal. This means that there would be no difference between a strong appeal and a weak (or even meaningless) one. While authority seems to be a useful factor in evaluating arguments in an organisation, it seems unreasonable to rely solely on it. There are, therefore, opportunities for combining authority with the other argument evaluation techniques described earlier.

4.1.2 Challenges

The discussion above shows that the nature of argument evaluation depends largely on the object of negotiation and the way that agents represent and update their internal mental states. For example, in the framework presented by Parsons *et al.* (1998), agents are able to perform some objective argumentation over their beliefs about the availability of resources, the achievability of intentions, and so on. This allows agents to potentially modify each other's mental attitudes, which may influence their preferences. In frameworks such as those of Ramchurn *et al.* (2003b) and Kraus *et al.* (1998), on the other hand, evaluation is based solely on the direct comparison of expected utilities. Agents do not influence each other's beliefs, but rather exert pressure on each other by exercising their ability to influence the outcomes (for example, by making a promise or a threat). In other words, an agent would not *voluntarily modify* its position as a result of correcting its perceptions of the environment, but rather *forcedly concede* on its position as a result of pressure from its counterpart. Many opportunities exist for combining the objective (belief-based) and subjective (value-based) approaches to argument evaluation. For example, how can we combine the objective evaluation of the logical form of an argument with a subjective evaluation of its consequences based on utility, trust, authority, etc.?

Another challenge is that of providing unified argumentation frameworks that facilitate negotiation dialogues involving the notions of goals, beliefs, plans, etc. Rahwan *et al.* (2003c) argued that systems of argumentation designed for arguing about beliefs are not readily suitable for allowing for argumentation over goals, particularly due to the different ways conflict resolution among arguments must be dealt with. For example, there is a difference between attacking a goal by demonstrating that it is not achievable and attacking it by demonstrating that it is not useful.

Rahwan *et al.* (2003b, c) demonstrated different ways in which goals may relate to their sub-goals, their super-goals and the agent's beliefs. This allows one to characterise different types of arguments that may be provided against a particular goal, and how they can, if successful, affect the agents' mental states. For example, the following statement represents an argument stating that the goal of going to Sydney is *justified* by the belief that there is a conference there (written *justify*(*confInSyd*, *goSyd*)), that this goal is *achievable* by buying a ticket and arranging accommodation (written *achieve*({*buyTicket*, *arrangeAccomm*}, *goSyd*)), and that it is *instrumental* towards the more basic goal of presenting a paper, (written *instr*(*goSyd*, *presentPaper*)):

$$\langle(\{presentPaper\}, \{confInSyd\}, \{buyTicket\}, arrangeAccomm\}: goSyd\rangle$$

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

The following are some ways of attacking the argument for the goal *goSyd*.

1. **Attack:** present statement $\neg$ *achieve*({*buyTicket*, *arrangeAccomm*}, *goSyd*). Here, the counterpart attacks the relation between the subgoals and the goal in question by arguing that buying a ticket and arranging accommodation are not sufficient for going to Sydney (say, one also needs to book a taxi to the airport).
   Effect: the statement $\neg$ *achieve*({*buyTicket*, *arrangeAccomm*}, *goSyd*) is removed from the knowledge base. If no alternative plan is found, the goal is deemed unachievable and must be dropped.
2. **Attack:** present statement *instr*(*goPerth*, *presentPaper*). Here, the counterpart provides an alternative plan for presenting the paper without having to go to Sydney (say, there is a similar conference in Perth).
   Effect: statement *instr*(*goPerth*, *presentPaper*) is added to the agent's knowledge base. The agent then compares the proposed plan (involving going to Perth) with the existing plan (involving going to Sydney), and based on the outcome of this comparison, the goal *goSyd* might be dropped (along with the whole plan to which it belongs).

The paper discusses other types of possible attacks, as a preliminary step to understanding the *space of possible influences* ABN agents may (or must be able to) exert in the course of dialogue.

Another attempt to unify argumentation over beliefs with argumentation over values was presented by Fox and Parsons (1998). In this framework, Fox and Parsons distinguished between *belief arguments* and *value arguments*. A value argument represents the value of a state or condition from a particular point of view. The authors provide a qualitative probabilistic model for *flattening* arguments relating to the same conclusion, and *combining* arguments relating to different conclusions to make composite conclusions. The belief-argument flattening and combination functions satisfy the axioms of probabilistic reasoning about beliefs, while the value-argument functions satisfy the axioms of utility theory. An agent can then use a third pair of flattening and combination functions to derive the *expected value* of the situation(s) resulting from executing an action. This involves considering the values of resulting states as well as the probabilities of their occurrence. Therefore, this can be seen as an argumentation-based qualitative decision theory.

4.2 Argument and proposal generation

Another central problem in the study of argumentation is that of argument generation. This problem is concerned with generating *candidate* arguments[24] to present to a dialogue counterpart. These arguments are usually sent in order to entice the counterpart to accept some proposed agreement. Hence, in negotiation, argument and proposal generation are closely related processes.

4.2.1 State of the art

In existing ABN frameworks, proposal generation is usually made as a result of some utility evaluation or planning process. For example, Sierra *et al.* (1998) and Ramchurn *et al.* (2003b) assumed that agents have a means of generating proposals that increase (or maximise) their utilities. Kraus *et al.* (1998), Parsons *et al.* (1998), and Sadri *et al.* (2001b) used an underlying planner to generate a set of actions or resources needed to achieve some intention. Agents then request the actions or resources they cannot achieve or obtain on their own from other agents. If they fail to obtain immediate acceptance, they may propose to perform an action (or set of actions) or to provide resources in return for acceptance. This may be done by just giving away what they do not need, or by measuring the utilities they lose in the exchange.

Proposals may be accompanied by arguments. In the framework of Kraus *et al.* (1998), for example, agents may choose to accompany proposals with arguments generated using explicit rules. By means of an illustration, the following is an informal description of the threat-generation rule for agent *i*:

[24] We leave the discussion of selecting the *best* argument to Section 4.3.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

IF
A request has been sent to agent j to perform action α &
j rejected this request &
j has goals g_1 and g_2 &
j prefers g_2 to g_1 &
doing α achieves $\neg g_1$ &
doing β achieves $\neg g_2$ &
i believes doing β is credible and appropriate
THEN
i requests α again with the following threat:
if you don't do α, I will do β

If the rule body is satisfied, the corresponding threat will become a *candidate argument*. The agent may generate other candidate arguments, such as promises or appeals, using other rules.

In a similar fashion, Ramchurn *et al.* (2003b) provided ''preconditions'' for each argument to become a candidate argument. For example, for an agent to promise to perform action α in return for β, it must believe the counterpart actually wants α to be executed, that its gain from having α executed outweighs the cost of it performing β, and so on.

As mentioned above, the frameworks of Parsons *et al.* (1998), Sadri *et al.* (2001b) and Amgoud *et al.* (2000b) take a planning approach to proposal generation. Arguments are, in fact, generated in the process of proposal generation itself. In other words, an agent justifies a request by simply stating the truth about its needs, plans, underlying assumptions, and so on, which ultimately caused the need to arise. This is different from the other utility-based approaches described above, where agents can, in a sense, *create* arguments, such as threats and rewards, by exploiting their abilities to influence the outcomes. Of course, there is nothing that directly prevents agents from combining the two.

As described earlier, Rahwan *et al.* (2003b) provided a characterisation of the types of arguments an agent can make in relation to the goal and belief structures of its counterpart. This provides a more fine-grained portfolio of candidate arguments than those of Parsons *et al.* (1998), Sadri *et al.* (2001b) and Amgoud *et al.* (2000b) (where only plans or promises can be put forward as arguments).

Finally, authority could also be used in argument generation. Sierra *et al.* (1998), for example, defined a simple authoritarian agent, which always exploits its social power by threatening whenever possible.

4.2.2 Challenges

More work needs to be done in order to provide a unified way of generating arguments by considering both objective and subjective criteria. Moreover, there is a need for a complete characterisation of the *space* of possible arguments. This is not necessarily a trivial task since in some frameworks the number of possible arguments may be infinite (say, if the framework allows for nested arguments about what may happen in the future or nested dialogues).

More work is also needed to understand the influence of different factors, such as the interaction protocol, authority, expected utility, honesty, etc., on argument generation. Specifically, how can authority be used in constructing an argument? Should an agent believe in an argument in order to present it? Can agents bluff? These are a few of the questions that need to be answered before a complete framework for argument generation is achieved.

4.3 Argument selection

An issue related to the problem of argument generation is that of argument selection. The question of argument selection is as follows: given a number of candidate arguments that an agent may utter to its counterpart, which is the ''best'' argument from the point of view of the speaker?

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

Note that argument selection may take place in conjunction with argument generation. An agent need not generate all possible arguments before it makes a selection of the most suitable one. Instead, the agent may only concern itself with the generation of the most suitable argument itself. In other words, the agent might prune the set of candidate arguments during the process of argument generation. Whether or not this is possible, of course, depends on the nature of the argumentation framework underlying the agent's decision making.

4.3.1 State of the art

In the work of Kraus *et al.* (1998), arguments are selected according to the following argument strength order, with threats being the strongest arguments.

1. Appeal to prevailing practice.
2. A counter example.
3. An appeal to past promise.
4. An appeal to self-interest.
5. A promise of future reward.
6. A threat.

The intuition is that a negotiator would progress from weak arguments up to the strongest. For example, there is no need to threaten the counterpart if an appeal is sufficient to persuade them to accept a request. The authors argue that generating appeals is less costly to the persuader than threats or rewards since the latter involve possible negative side-effects.

In Ramchurn *et al*'s framework (Ramchurn *et al.*, 2003b), agents factor trust and utility in order to decide which candidate argument to send with a request. The following example rule is provided.

`Rule 1`: *if* ***trust*** *is* `low` *and* ***utility*** *of the proposal is* `high`
(*I need to do X and I don't trust you*)
then send a ***strong argument***
`Rule 2`: *if* ***trust*** *is* `high` *and* ***utility*** *of the proposal is* `low`
(*I don't really need to do X and I trust you*)
then send a ***weak argument***

In this rule, `low` and `high` are linguistic variables manipulated using fuzzy operators. The stronger an argument is, the more it is likely to lessen the opponent's trust in the proponent and the more it could coerce the opponent to change its preferences (e.g., by making a significant threat). However, this lowering of trust results in less cooperative behaviour which, in turn, makes it harder for the proponent to persuade the opponent to accept its future proposals. Thus strong arguments should only be sent when the negotiation needs to take place in the shortest time possible, when the proposal has high utility for the proponent or when it is known that the other partner cannot be trusted to reach effective agreements efficiently. Otherwise weaker arguments should be used[25].

In other frameworks, argument generation is based on the relationships between arguments. Agents in the framework presented by Parsons *et al.* (1998) provide the strongest argument possible based on the acceptability classes (e.g., a tautological argument if possible). For Amgoud *et al.* (2000b), agents compare arguments based on preferential ordering over their constituent propositions in a similar manner to that in argument evaluation (i.e. based on the argumentation system of Dung (1995)). Finally, for Sadri *et al.* (2001b), agents can compare the costs of different alternative plans to present to the counterpart.

4.3.2 Challenges

The problem of argument selection can be considered as the essence of *strategy* in ABN dialogues in general (provided the candidate arguments contain all possible arguments). However, there is

[25] Refer to the original paper for some related empirical observations.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

very little existing work on strategies in multi-agent dialogues. Some work is emerging that investigates strategic move selection in persuasion dialogues (Amgoud & Maudet, 2002), as well as in inquiry and information-seeking dialogues (Parsons *et al*., 2002, 2003). Similar work needs to be done on ABN dialogues in order to provide a sound theoretical base for potential applications. Rahwan *et al*. (2003a) provide a preliminary, informal attempt at characterising strategic factors in negotiation dialogues. In this work, strategies depend on various factors, such as the agents' goals, the interaction protocol, the agents' capabilities, the resources available to participants, and so on.

Suitable argument selection in a negotiation context must take into account information about the negotiation counterpart. To this end, in Bayesian game theory (Osborne & Rubinstein, 1994), a counterpart is modelled by a probability distribution representing the uncertainty of the first party regarding the counterparts' initial information and the payoffs they receive from the different outcomes (i.e. action profiles). This raises the opportunity to use *learning* techniques in order to find patterns in the counterpart's behaviour and use these findings in future encounters with the same (or similar) counterpart(s). Sandholm and Crites (1995), for example, apply reinforcement learning in the context of the iterated Prisoner's Dilemma game to allow agents to better predict the patterns of behaviour of their opponents. Bayesian learning in less-restricted negotiation protocols has also been investigated by Zeng and Sycara (1997). In ABN, more sophisticated models of the negotiation counterparts are needed, and appropriate methods of updating these models are essential for understanding the dynamics of opponents' strategies, preferences, beliefs, etc. This is a particularly challenging task for ABN since agents may not only model the observed ''behaviour'' of one another, but also the ''mental attitudes'' motivating that behaviour. Another important question is whether and how such learning agents converge to better and quicker deals in multiple negotiation encounters.

5 Summary of the ABN state of the art

Having analysed the various frameworks in detail in the previous sections, we now proceed to present a high-level view of what has been achieved in the field of ABN as a whole. In this way, we aim to identify the areas that have been extensively researched, the key results in those areas and those areas that have not received sufficient attention from the research community.

In Table 2, we compare the different existing frameworks in terms of their main characteristics[26]. Specifically, the first column describes the style of argumentation underlying the ABN framework. This covers the informal literature that motivates and provides intuitive backing of the research, as well as the formal theories underlying the specification of the framework (e.g., decision theory, argumentation theory, dialogue games, etc.). When taken together this provides an idea of the starting point of each framework. As can be seen, frameworks such as those of Amgoud *et al*. (2000b) and Sadri *et al*. (2001b) start from a single-agent proof procedure and try to split it into multiple disjoint agents while preserving the correctness of the proof theory. In particular, Amgoud *et al*. (2000b) build on an existing framework for belief-based argumentation in a single agent (Amgoud & Cayrol, 1998), and view the proof theory of that framework as a dialogue between two agents. Sadri *et al*. (2001b) take a similar approach by providing a dialectical version of an abductive logic programming framework (Fung & Kowalski, 1997). In contrast, frameworks such as those presented by McBurney *et al*. (2003) and Rahwan *et al*. (2003b) start by discussing the different types of interaction patterns needed among agents, and from there attempt to create a dialogue system. The former approach has the advantage of being built on solid theories that make the interaction, in a sense, more predictable. On the other hand, the latter approach tends not to specify the internal decision mechanisms of the agents, and concentrates instead on studying the general properties of the dialogue itself or the interesting types of influences that agents might be able to exert on one another. We believe that as the field of ABN matures, these approaches will

[26] Wherever the framework in question has not addressed the particular attribute of the table (e.g. Protocol) significantly, we note this as N/A.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

Table 2 Characteristics of extant ABN models

Authors	Characteristic			
	Argumentation style	Protocol	Main assumptions	Implementation
Kraus, Sycara and Evenchik	Based on the psychology of persuasion (Karlins & Abelson, 1970) Use threats, rewards and appeals	Implicit in agent specification	Agents have utility function Agents share same architecture	Blocks World Scenario (implemented)
Sierra, Jennings, Noriega & Parsons	Same as above	Finite-state machine Allows generic meta information to be passed	Argument generation, selection and evaluation are predefined E-Institution present	N/A
Ramchurn, Jennings and Sierra	Same as above Framework based on Sierra *et al.* (1998)	N/A	States of the world have a known value Trust model implicit	Abstract world (simulated)
Parsons, Jennings and Sierra	Logic-based, inspired by Elvang-Gøransson *et al.* (1993a)	Finite-state machine Allows generic meta information to be passed	Agents are cooperative Agents share same architecture	N/A
Amgoud, Maudet and Parsons	Classification of dialogues based on Walton and Krabbe (1995) Logic-based	Dialogue game protocol, allows promises, challenges, assertions and requests to be sent	Agents share same architecture Complete preferences over knowledge bases Agents share preferences	N/A
Sadri, Torroni and Toni	Based on abductive logic programming View of dialogue from Walton and Krabbe (1995)	Implicit in agent specification	Agents share same architecture Agents are cooperative	Logic agents negotiate to exchange tools
McBurney, van Eijk, Parsons and Amgoud	View of negotiation based on Walton and Krabbe (1995) Agent model influenced by consumer modelling models in marketing (Lilien *et al.*, 1992)	Dialogue game protocol allows potential offers and preference statements to be passed	Agents do not decide as a group, goods purchased afford negotiation Agent has utility function and other agents might know its valuation for goods	N/A
Rahwan, Sonenberg and Dignum	Approach influenced by interest-based negotiation, among humans (Fisher & Ury, 1983)	Sketched locutions	Agents have comparison criteria for selecting goals based on their support	N/A

meet in the middle, achieving rich interactions on the system level, as well as comprehensive and verifiable proof theories on the detailed level.

The next column describes the protocol. It is clear that some ABN frameworks have not yet addressed the protocol definition, while in others it is the mainstay of their contribution. Moreover, the frameworks can differ in the way they specify the protocols, by making them implicit or explicit, defining them as finite-state machines, as dialogue games, and so on. The third column describes some of the important assumptions that each framework makes. In some frameworks, for instance, the agents must be cooperative for ABN to work. Frameworks can also vary in their assumptions about agents' utilities and preferences. Finally, we have specified whether the framework has been implemented and, if so, what form this takes.

In Table 3, we outline the various frameworks in terms of whether and how each framework addresses the problems of argument generation, selection, and evaluation. One important observation from this is that argument selection has had very little attention in the ABN community. This is, we believe, partly because effective strategies for deciding what arguments to utter are likely to be protocol-dependent. There is still no formal theory of interaction protocols covering all types of mechanism. It is to be expected, therefore, that such work will focus first on defining the protocols and exploring their properties, rather than on devising strategies for participants using the protocols.

As can be seen, there is a clear contrast in the way the three main mechanisms are conceived by the different frameworks. We contend that this is mainly due to the differences in the underlying style of argumentation. However, despite these differences, their contributions are broadly complementary.

In summary, the frameworks reviewed in this paper represent different preliminary attempts at solving parts of the puzzle by: (i) constructing generic models of ABN (Sierra *et al.*, 1998); (ii) constructing limited, yet implementable systems and studying their applicability (Kraus *et al.*, 1998; Sadri *et al.*, 2001b); (iii) studying the applicability of particular logic-based argumentation frameworks to ABN (Parsons *et al.*, 1998; Amgoud *et al.*, 2000a); (iv) studying the properties of different decision making components and concepts such as trust in controlled settings (Ramchurn *et al.*, 2003b); (v) constructing specific complete negotiation protocols necessary for facilitating particular types of ABN (McBurney *et al.*, 2003); and (vi) studying the different types of influences that can be attempted by participants in an ABN dialogue (Rahwan *et al.*, 2003b, c).

6 Conclusions and future directions

Argumentation is gaining increasing importance as a fundamental concept in multi-agent interaction, mainly because it enables rational dialogue (McBurney, 2002) – the giving and receiving of reasons for statements – and because it enables richer forms of negotiation than have hitherto been possible in game-theoretic or heuristic-based models. Against this background, this paper has sketched the landscape of the emerging research field of ABN and reviewed the state of the art. Specifically, we have identified the key advantages that argumentation adds to existing models of negotiation. We then provided a sketch of the different features and decision components needed to facilitate ABN and used these to describe and analyse a variety of existing ABN frameworks. Even though the research area is still in its infancy, many important lessons have been learned through the various attempts thus far. We now have insights about the various functionalities needed for ABN, some possible ways to achieve these functionalities, and, most importantly, major challenges and open questions in the field.

This analysis has highlighted the fact that there is clearly a need for more work on identifying different ways of implementing mechanisms for argument evaluation, generation and selection. In particular, this involves understanding how these processes are related to the agent's underlying notions of rationality. In other words, there is a need for a better understanding of how agents may use objective argument-based reasoning about the state of the world to achieve

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

Table 3 Necessary mechanisms for an ABN framework

Authors	Mechanism		
	Generation	Selection	Evaluation
Kraus, Sycara and Evenchik	Rules determine possible arguments from available pool	Rigid order among argument *classes* only (start with appeals then rewards, then threats)	First using rules, then taking the conflicting intentions into account
Sierra, Jennings, Noriega and Parsons	Partially specified by roles	N/A	Only rules based on authority are presented
Ramchurn, Jennings and Sierra	Possible arguments chosen from pool according to preconditions	Taking into account utility of proposal and trust to choose argument of particular strength	According to utility gained and trust
Parsons, Jennings and Sierra	Starting from existing intentions, generate tentative plans and proofs (arguments) using simple BDI planning rules	Possibly choose strongest argument based on acceptability classes	Based on acceptability classes Accepted if argument cannot be logically defeated
Amgoud, Maudet and Parsons	N/A	N/A	Based on interaction between arguments and preferences over their contents
Sadri, Torroni and Toni	Based on rules	N/A	Compare plans with or without exchange of resources
McBurney, van Eijk, Parsons and Amgoud	N/A	N/A	Informally specified
Rahwan, Sonenberg and Dignum	Based on a list of possible attacks on goals	N/A	Based on the effect of updating relationships between goals, supergoals, subgoals and beliefs

subjective goals, such as maximising utility, satisfying preferences, fulfilling social norms, and so on.

This raises fundamental questions about the relationship between argumentation-based approaches to group decision making and the game-theoretic approaches dictated by the traditional economic conception of rationality. Young (2001) argues that the traditional notion of economic rationality fails to capture the process by which agents form their preferences during negotiation. In Young's words:

> ''. . . since, under the assumptions of the received concept of economic rationality, each player's objectives or ends are not justifiable rationally, and are set by the agent's preferences in advance of reasoning, game theory has no way to capture or evaluate those ends. It can never be sure it understands the complete motivational set which is driving the individual, and in the end can only reduce the rich mix of factors which motivate and guide human beings to simple economic self-interest. Unfortunately, . . . much is lost along the way.', (Young, 2001, p. 97.)

Young then argues that in order to solve this problem, agents must adopt a notion of *communicative rationality* rather than merely strategic, instrumental rationality. We take the position that argumentation, which allows agents to critically evaluate their and each others' underlying motivational attitudes during negotiation, can enable agents to realise deals not possible by following the game-theoretic model of negotiation. However, the exact ''formal'' relationship between ABN and game-theoretic models of strategic decision making remains a fertile area of research.

Another important challenge facing future research is understanding the ''social'' aspects of ABN in societies of agents. There is some existing work on investigating the effects of norm adoption (Glass & Grosz, 2003; Castelfranchi *et al*., 1998) and social influence (Panzarasa & Jennings, 2002; Sen *et al*., 2002) on decision making from both the individual and collective agent perspectives. However, there is no generic formal theory that establishes a precise relationship between normative social behaviour and the resulting outcomes of communication. A full investigation of these issues will also need to respond to the *theory of communicative action* of Habermas (1984), and the associated philosophical issues concerning group decision making.

Another important social aspect is ''trust'', which plays a crucial role when agents cannot be assumed to keep to their commitments. Trust enables the selection of the most appropriate negotiation partners and affects the way interaction unfolds. It also enables agents to improve the outcome in repeated encounters. For example, in iterated game-theory, trust has proven useful in allowing agents to converge to higher-payoff Nash equilibria (Mukherjee *et al*., 2001). There are opportunities for using trust models in ABN, over single and repeated encounters, in order to guide the selection of negotiation strategies. In the context of ABN, Ramchurn *et al*. (2003b) used trust in argument evaluation and generation in an ABN setting. A more elaborate model of trust could also be incorporated (e.g., Ramchurn *et al*., 2003a; Sabater & Sierra, 2002). In the more general context of argumentation, Parsons and Giorgini (2000) explored the evaluation of the strength of arguments based on the reliability of the agents providing these arguments.

One of the significantly unexplored topics is *mediated* negotiation, in which a trusted, neutral third party assists negotiators in reconciling their views and interests. Several multi-party negotiation systems with artificial mediators have been built to aid human negotiation. For example, the PERSUADER system (Sycara, 1985, 1992) is a mediating system that relies on the construction of hierarchical goal structures for participants and attempts to reconcile these. However, from the descriptions given, it is not apparent how the framework, which mostly describes the reasoning mechanism for a single agent mediating between human opponents, could be extended to (artificial) multi-agent systems. In another example, the ZENO system (Gordon & Karacapilidis, 1997; Gordon *et al*., 1997) permits human participants to undertake joint deliberation about urban planning decisions, and provides an intelligent assistant to a human

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

mediator. Major issues for these systems concern the nature of the mediator (is it a human or a software agent?) and its relationship to the negotiators (does it merely assist them in identifying new proposals or can it enforce its will on them?). What is still absent in this area is a grounded theory of mediation in a multi-agent negotiation.

Understanding the computational complexity of the ABN process is important before ABN frameworks can be used in real-world applications. In this context, complexity may arise, for example, from:

- the internal mental processes of the participants (e.g., preference determination, argument assessment, strategy generation, strategy selection, argument generation, argument selection, etc.);
- the assessment of uttered locutions by the agent platform to determine conformance with the protocol participants;
- the achievement of particular negotiation outcomes (e.g., complexity of successful termination).

Some work has been done on aspects of ABN complexity. For example, relative to the assessment of uttered illocutions and conformance to protocols, Wooldridge and Parsons (2000) studied complexity issues in logic-based negotiation protocols. More recently, with regards to the achievement of dialogue outcomes, Parsons *et al.* (2002) studied the termination and complexity of information seeking, inquiry and persuasion dialogues.

In addition to the work reported in this survey, there have been attempts to study the process of ABN from a higher level of abstraction and analyse some of the convergent properties associated with certain assumptions (very much like the work in game theoretic models discussed in Section 2.1). For example, Tohmé (1997) views negotiation as resource allocation with uncertainty caused by imperfect information about others. The agents exchange messages that correspond to updates of beliefs and consequently cause new messages to be generated. Tohmé showed that under certain conditions, beliefs converge in the long run, leading to successful negotiation (i.e. agreement). He also showed, again under certain conditions, that such agreement could take place even though no interaction protocol has been defined. There is a need for further studies in this direction in order to understand the general dynamics of ABN systems without necessarily subscribing to particular internal computational mechanisms within the agents.

In conclusion, we see ABN as an important and challenging research area that combines the study of agent architectures, decision making, relationships and interaction from computer science, economics and organisation theory, with the study of argumentation in dialogue, logic and psychology. Existing work has begun to address different aspects of the challenge, but much remains to be done, on both the conceptual and technical levels. ABN will enable us to build more sophisticated, flexible and robust negotiating agents, capable of operating in more dynamic, uncertain and unpredictable environments. Moreover, research into ABN has the potential to inform our understanding of strategic interaction among humans, and hence to contribute to related fields.

Acknowledgements

During this work, IR was supported by a Melbourne Research Scholarship (MRS) and a Postgraduate Overseas Research Experience Scholarship (PORES) from the University of Melbourne, and a PhD top-up scholarship from CSIRO Mathematical and Information Sciences, Australia. SDR is supported by the FEEL project (IST-2000-26135) funded by the EU Commission through the Disappearing Computer Initiative. This work commenced while IR was a visitor to the IAM Group of the School of Electronics and Computer Science at the University of Southampton, and the Agent ART Group of the Department of Computer Science at the University of Liverpool, UK. This research was also partially supported by NSF-REC-02-19347 and NSF-IIS-0329037.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

References

Amgoud, L. & Cayrol, C. 1998 On the acceptability of arguments in preference-based argumentation. In *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI 1998), San Francisco, CA*. San Francisco, CA: Morgan Kaufmann, pp. 1–7.

Amgoud, L. & Maudet, N. 2002 Strategical considerations for argumentative agents (preliminary report). In Benferhat, S. & Giunchiglia, E. (eds.), *Proceedings of the 9th International Workshop on Non-Monotonic Reasoning (NMR 2002): Special Session on Argument, Dialogue and Decision*, pp. 399–407.

Amgoud, L., Maudet, N. & Parsons, S. 2000a Modelling dialogues using argumentation. In Durfee, E. (ed.), *Proceedings of the 4th International Conference on Multi-Agent Systems (ICMAS 1998), Boston, MA*. IEEE Press, pp. 31–38.

Amgoud, L., Maudet, N. & Parsons, S. 2002 An argumentation-based semantics for agent communication languages. In Harmelen, F. V. (ed.), *Proceedings of the European Conference on Artificial Intelligence (ECAI-2002), Lyon, France*. IOS Press, pp. 38–42.

Amgoud, L. & Parsons, S. 2001 Agent dialogues with conflicting preferences. In Meyer, J.-J. & Tambe, M. (eds.), *Proceedings of the 8th International Workshop on Agent Theories Architectures and Languages (ATAL 2001), Seattle*, pp. 190–205.

Amgoud, L., Parsons, S. & Maudet, N. 2000b Arguments, dialogue, and negotiation. In Horn, W. (ed.), *Proceedings of the European Conference on Artificial Intelligence (ECAI 2000), Amsterdam, Netherlands*. IOS Press, pp. 338–342.

Aristotle. 1928 In Ross, W. D. (ed.), *Topics*. Oxford: Clarendon.

Ashri, R., Rahwan, I. & Luck, M. 2003 Architectures for negotiating agents. In Marik, V., Müller, J. & Pechoucek, M. (eds.), *Multi-Agent Systems and Applications III: Proceedings of the 3rd International Central and Eastern European Conference on Multi-Agent Systems (CEEMAS 2003) (Lecture Notes in Artificial Intelligence, Vol. 2691)*. Berlin: Springer-Verlag, pp. 136–146.

Austin, J. L. 1962 *How To Do Things with Words*. Oxford: Oxford University Press. (Originally delivered as the William James Lectures at Harvard University in 1955.)

Bartolini, C., Preist, C. & Jennings, N. R. 2002 Architecting for reuse: a software framework for automated negotiation. In Giunchiglia, F., Odell, J. & Weiss, G. (eds.), *Proceedings of the 3rd International Workshop on Agent-Oriented Software Engineering (AOSE-2002)*. Berlin: Springer-Verlag.

Bench-Capon, T. J. M. 2001 Truth and consequence: complementing logic with value in legal reasoning. *Information and Communications Technology Law* **10**(1).

Castelfranchi, C., Conte, R. & Paolucci, M. Normative reputation and the costs of compliance. *Journal of Artificial Societies and Social Simulation*, 1998.

Chesñevar, C. I., Maguitman, A. & Loui, R. 2000 Logical models of argument. *ACM Computing Surveys* **32**(4), 337–383.

Coleman, J. S. 1990 *Foundations of Social Theory*. Cambridge, MA: Harvard University Press.

Colombetti, M. 2000 A commitment-based approach to agent speech acts and conversations. In *Proceedings of the Workshop on Agent Languages and Conversational Policies*, pp. 21–29.

Conitzer, V. & Sandholm, T. 2002 Complexity of mechanism design. In Darwiche, A. & Friedman, N. (eds.), *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence (UAI-2002)*, pp. 103–110.

Dash, R. K., Parkes, D. C. & Jennings, N. R. 2003 Computational mechanism design: a call to arms. *IEEE Intelligent Systems* **18**(6), 40–47.

de Boer, F., van Eijk, R. M., van der Hoek, W. & Meyer, J.-J. 2003 A fully abstract model for the exchange of information in multi-agent systems. *Theoretical Computer Science* **290**(3), 1753–1773.

Dignum, F., Dunin-Kęplicz, B. & Berbrugge, R. 2000 Agent theory for team formation by dialogue. In Castelfranchi, C. & Lespérance, Y. (eds.), *Intelligent Agents VII: Proceedings of the 7th International Workshop on Agent Theories, Architectures, and Languages (ATAL-2000) (Lecture Notes in Computer Science, Vol. 1986)*. Berlin: Springer-Verlag, pp. 150–166.

Dung, P. M. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and *n*-person games. *Artificial Intelligence* **77**(2), 321–358.

Durfee, E. H. 1999 Practically coordinating. *Artificial Intelligence Magazine* **20**(1), 99–116.

Elvang-Gøransson, M., Krause, P. & Fox, J. 1993a Acceptability of arguments as ''logical uncertainty''. In Clarke, M., Kruse, R. & Moral, S. (eds.), *Proceedings of the European Conference on Symbolic and Quantitative Approaches to Reasoning and Uncertainty (ECSQARU 1993) (Lecture Notes in Computer Science, Vol. 747)*. Berlin: Springer-Verlag, pp. 85–90.

Elvang-Gøransson, M., Krause, P. & Fox, J. 1993b Dialectic reasoning with inconsistent information. In Heckerman, D. & Mamdani, A. (eds.), *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence, Washington, DC*. Morgan Kaufmann, pp. 114–121.

Esteva, M., Rodríguez, J. A., Sierra, C., Garcia, P. & Arcos, J. L. 2001 On the formal specifications of electronic institutions. In Dignum, F. & Sierra, C. (eds.), *Agent Mediated Electronic Commerce*

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

(The European AgentLink Perspective) (Lecture Notes in Artificial Intelligence, Vol. 1991). Berlin: Springer-Verlag, pp. 126–147.

Faratin, P. 2000 *Automated Service Negotiation Between Autonomous Computational Agents*. PhD thesis. Department of Electronic Engineering, Queen Mary and Westfield College, University of London.

Faratin, P., Sierra, C. & Jennings, N. R. 1998 Negotiation decision functions for autonomous agents. *International Journal of Robotics and Autonomous Systems* **24**(3–4), 159–182.

Faratin, P., Sierra, C. & Jennings, N. R. 2002 Using similarity criteria to make trade-offs in automated negotiations. *Artificial Intelligence* **142**(2), 205–237.

Fatima, S., Wooldridge, M. & Jennings, N. R. 2001 Optimal negotiation strategies for agents with incomplete information. In Meyer, J.-J. & Tambe, M. (eds.), *Intelligent Agent Series VIII: Proceedings of the 8th International Workshop on Agent Theories, Architectures, and Languages (ATAL 2001) (Lecture Notes in Computer Science, Vol. 2333)*. Berlin: Springer-Verlag, pages 53–68.

Fatima, S., Wooldridge, M. & Jennings, N. R. 2002 Multi-issue negotiation under time constraints. In Castelfranchi, C. & Johnson, L. (eds.), *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2002), New York, USA*. ACM Press, pp. 143–150.

Fatima, S., Wooldridge, M. & Jennings, N. R. 2004 An agenda based framework for multi-issues negotiation. *Artificial Intelligence Journal* **152**(1), 1–45.

FIPA. 2001 Communicative Act Library Specification. 2001 Technical report XC00037H, Foundation for Intelligent Physical Agents.

Fisher, R. & Ury, W. 1983 *Getting to Yes: Negotiating Agreement Without Giving In*. New York: Penguin Books.

Fox, J. & Parsons, S. 1998 Arguing about beliefs and actions. In Hunter A. & Parsons, S. (eds.), *Applications of Uncertainty Formalisms (Lecture Notes in Computer Science, Vol. 1455)*. Berlin: Springer-Verlag, pp. 266–302.

Fung, T. & Kowalski, R. 1997 The iff proof procedure for abductive logic programming. *Journal of Logic Programming* **33**(1), 151–165.

Glass, A. & Grosz, B. 2003 Socially conscious decision-making. *Autonomous Agents and Multi-Agent Systems* **6**(3), 317–339.

Gordon, T. F. & Karacapilidis, N. 1997 The Zeno argumentation framework. In *Proceedings of the Sixth International Conference on AI and Law, New York, USA*. ACM Press, pp. 10–18.

Gordon, T. F., Karacapilidis, N., Voss, H. & Zauke, A. 1997 Computer-mediated cooperative spatial planning. In Timmermans, H. (ed.), *Decision Support Systems in Urban Planning*. E and F. N. Spon, pp. 299–309.

Habermas, J. 1984 *The Theory of Communicative Action, Volume 1: Reason and the Rationalization of Society*. Boston, MA: Beacon Press.

Hamblin, C. L. 1970 *Fallacies*. London: Methuen.

Harsanyi, J. C. 1956 Approaches to the bargaining problem before and after the theory of games: a critical discussion of Zeuthen's, Hicks', and Nash's theories. *Econometrica* **24**, 144–157.

Hendler, J. & McGuinness, D. L. 2000 The DARPA agent markup language. *IEEE Intelligent Systems* **15**(6), 67–73.

Hitchcock, D. 2002 Aristotle's theory of argument evaluation. In Boudouris, K. & Poulakos, J. (eds.), *The Philosophy of Communication*, Vol. 1. Athens: Ionia Press.

Huget, M.-P. & Wooldridge, M. 2003 Model checking for ACL compliance verification. In *Proceedings of the Workshop on Agent Communication Languages and Conversation Policies*. To appear.

Jennings, N. R. 2000 On agent-based software engineering. *Artificial Intelligence* **117**(2), 277–296.

Jennings, N. R. 2001 An agent-based approach for building complex software systems. *Communications of the ACM* **44**(4), 35–41.

Jennings, N. R., Faratin, P., Lomuscio, A. R., Parsons, S., Sierra, C. & Wooldridge, M. 2001 Automated negotiation: prospects, methods and challenges. *International Journal of Group Decision and Negotiation* **10**(2), 199–215.

Jennings, N. R., Parsons, S., Noriega, P. & Sierra, C. 1998 On argumentation-based negotiation. In *Proceedings of the International Workshop on Multi-Agent Systems, Boston*, pp. 1–7.

Kakas, A. & Moraitis, P. 2003 Argumentation based decision making for autonomous agents. In *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2003), Melbourne, Australia*, pp. 883–890.

Karlins, M. & Abelson, H. I. 1970 *Persuasion: How Opinions and Attitudes are Changed*, 2nd edn. Lockwood.

Kowalczyk, R. 2000 On negotiation as a distributed fuzzy constraint satisfaction problem. In *Proceedings of the 3rd International Symposium on Soft Computing for Industry*. World Automation Congress, pp. 631–637.

Kowalczyk, R. & Bui, V. 2001 On constraint-based reasoning in e-negotiation agents. In Dighum, F. & Cortés, U. (eds.), *Agent-Mediated Electronic Commerce III (Lecture Notes in Computer Science, Vol. 2003)*. Berlin: Springer-Verlag, pp. 31–46.

Kraus, S. 2001 *Strategic Negotiation in Multi-Agent Environments*. Cambridge, MA: MIT Press.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

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

Kraus, S., Sycara, K. & Evenchik, A. 1998 Reaching agreements through argumentation: a logical model and implementation. *Artificial Intelligence* **104**(1–2), 1–69.

Labrou, Y., Finin, T. & Peng, Y. 1999 Agent communication languages: the current landscape. *Intelligent Systems* **14**(2), 45–52.

Larson, K. & Sandholm, T. 2002 An alternating offers bargaining model for computationally limited agents. In Castelfranchi, C. & Johnson, L. (eds.), *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2002)*, pp. 135–142.

Lilien, G. L., Kotler, P. & Moorthy, S. K. 1992 *Marketing Models*. Englewood Cliffs, NJ: Prentice-Hall.

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

Luo, X., Jennings, N. R., Shadbolt, N., Leung, H.-F. & Lee, J. H.-M. 2003 A fuzzy constraint-based knowledge model for bilateral, multi-issue negotiations in competitive environments. *Artificial Intelligence* **148**(1–2), 53–102.

Luo, X., Zhang, C. & Jennings, N. R. 2002 A hybrid model for sharing information between fuzzy, uncertain and default reasoning models in multi-agent systems. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems* **10**(4), 401–450.

Maudet, N. & Chaib-draa, B. 2003 Commitment-based and dialogue-game based protocols–new trends in agent communication language. *Knowledge Engineering Review* **17**(2), 157–179.

Mayfield, J., Labrou, Y. & Finin, T. 1996 Evaluating KQML as an agent communication language. In Wooldridge, M. J., Müller, J. P. & Tambe, M. (eds.), *Intelligent Agents II (Lecture Notes in Computer Science, Vol. 1039)*. Berlin: Springer-Verlag, pp. 347–360.

McBurney, P. 2002 *Rational Interaction*. PhD thesis, University of Liverpool.

McBurney, P., Parsons, S. & Wooldridge, M. 2002 Desiderata for agent argumentation protocols. In Castelfranchi, C. & Johnson, L. (eds.), *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2002)*, New York. ACM Press, pp. 402–409.

McBurney, P., van Eijk, R. M., Parsons, S. & Amgoud, L. 2003 A dialogue-game protocol for agent purchase negotiations. *Journal of Autonomous Agents and Multi-Agent Systems* **7**(3), 235–273.

McGuinness, D. L. 2001 Ontologies and online commerce. *IEEE Intelligent Systems* **16**(1), 8–14.

McGuinness, D. L. & van Harmelen, F. 2003 Web ontology language (OWL): overview. Technical report, W3C Working Draft.

Moulin, B. & Chaib-Draa, B. 1996 A review of distributed artificial intelligence. In O'Hare, G. & Jennings, N. R. (eds.), *Foundations of Distributed Artificial Intelligence*, John Wiley & Sons, pp. 3–55.

Mukherjee, R., Banerjee, B. & Sen, S. 2001 Learning mutual trust. In Falcone, R., Singh, M. P. & Tan, Y.-H. (eds.), *Trust in Cyber-societies (Lecture Notes in Computer Science, Vol. 2246)*. Berlin: Springer-Verlag, pp. 145–158.

Osborne, M. J. & Rubinstein, A. 1994 *A Course in Game Theory*. Cambridge, MA: MIT Press.

Panzarasa, P. & Jennings, N. R. 2002 Social influence, negotiation and cognition. *Simulation Modelling Practice and Theory* **10**(5–7), 417–453.

Panzarasa, P., Jennings, N. R. & Norman, T. J. 2002 Formalising collaborative decision making and practical reasoning in multi-agent systems. *Journal of Logic and Computation* **12**(1), 55–117.

Parsons, S. & Giorgini, P. 2000 An approach to using degrees of belief in BDI agents. In Bouchon-Meunier, B., Yager, R. R. & Zadeh, L. A. (eds.), *Information, Uncertainty and Fusion*. Dordrecht: Kluwer.

Parsons, S., Sierra, C. & Jennings, N. Agents that reason and negotiate by arguing. *Journal of Logic and Computation* **8**(3), 261–292.

Parsons, S., Wooldridge, M. & Amgoud, L. 2002 An analysis of formal inter-agent dialogues. In Castelfranchi, C. & Johnson, L. (eds.), *Proceedings of the 1st International Joint Conference on Autonomous Agents and MultiAgent Systems (AAMAS-2002), New York, USA*. ACM Press, pp. 394–401.

Parsons, S., Wooldridge, M. & Amgoud, L. 2003 On the outcomes of formal inter-agent dialogues. In Rosenschein, J., Sandholm, T., Wooldridge, M. & Yokoo, M. (eds.), *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2003)*. ACM Press, pp. 616–623.

Parunak, H. V. D. P. 1999 Industrial and practical applications of DAI. In Weiss, G. (ed.), *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*. Cambridge, MA: MIT Press, pp. 377–421.

Prakken, H. & Sartor, G. 2001 The role of logic in computational models of legal argument: a critical survey. In Kakas, A. & Sadri, F. (eds.), *Computational Logic: From Logic Programming into the Future (in honour of Bob Kowalski) (Lecture Notes in Computer Science, Vol. 2048)*. Berlin: Springer-Verlag, pp. 342–343.

Prakken, H. & Vreeswijk, G. 2002 Logics for defeasible argumentation. In Gabbay, D. & Guenthner, F. (eds.), *Handbook of Philosophical Logic*, Vol. 4, 2nd edn. Dordrecht: Kluwer, pp. 219–318.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

Pynadath, D. & Tambe, M. 2002 Multiagent teamwork: analyzing key teamwork theories and models. In Castelfranchi, C. & Johnson, L. (eds.), *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2002), New York, USA*. ACM Press, pp. 873–880.

Rahwan, I., Kowalczyk, R. & Pham, H. H. 2002 Intelligent agents for automated one-to-many e-commerce negotiation. In Oudshoorn, M. (ed.), *Proceedings of the 25th Australasian Conference on Computer Science*. Australian Computer Society Press, pp. 197–204.

Rahwan, I., McBurney, P. & Sonenberg, L. 2003a Towards a theory of negotiation strategy (a preliminary report). In Parsons, S. & Gmytrasiewicz, P. (eds.), *Proceedings of the 5th Workshop on Game Theoretic and Decision Theoretic Agents (GTDT-2003)*, pp. 73–80.

Rahwan, I., Sonenberg, L. & Dignum, F. 2003b On interest-based negotiation. In Dignum, F. (ed.), *Advances in Agent Communication (Lecture Notes in Artificial Intelligence, Vol. 2922)*. Berlin: Springer-Verlag.

Rahwan, I., Sonenberg, L. & Dignum, F. 2003c Towards interest-based negotiation. In Rosenschein, J., Sandholm, T., Wooldridge, M. & Yokoo, M. (eds.), *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2003)*. ACM Press, pp. 773–780.

Ramchurn, S., Sierra, C., Godo, L. & Jennings, N. R. 2003a A computational trust model for multi-agent interactions based on confidence and reputation. In *Proceedings of the workshop on Deception, Fraud and Trust in Agent Societies, Melbourne, Australia*, pp. 69–75.

Ramchurn, S. D., Jennings, N. R. & Sierra, C. 2003b Persuasive negotiation for autonomous agents: a rhetorical approach. In Reed, C., Grasso, F. & Carenini, G. (eds.), *Proceedings of the IJCAI Workshop on Computational Models of Natural Argument*. AAAI Press, pp. 9–17.

Rao, A. S. & Georgeff, M. P. 1995 Formal models and decision procedures for multi-agent systems. Technical report 61, Australian Artificial Intelligence Institute, Melbourne.

Rodriguez-Aguílar, J. A. & Sierra, C. 2002 Enabling open agent institutions. In Dautenhahn, K., Bond, A. H., Cañamero, L. & Edmonds, B. (eds.), *Socially Intelligent Agents: Creating Relationships with Computers and Robots*. Kluwer, pp. 259–266.

Rosenschein, J. & Zlotkin, G. 1994 *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers*. Cambridge, MA: MIT Press.

Rubiera, J. C., Lopez, J. M. M. & Muro, J. D. 2001 A fuzzy model of reputation in multi-agent systems. In *Proceedings of the 5th International Conference on Autonomous Agents*. ACM Press, pp. 25–26.

Rubinstein, A. 1997 *Modeling Bounded Rationality*. Cambridge, MA: MIT Press.

Sabater, J. & Sierra, C. 2002 Regret: a reputation model for gregarious societies. In Castelfranchi, C. & Johnson, L. (eds.), *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-2002)*. ACM Press, pp. 475–482.

Sadri, F., Toni, F. & Torroni, P. 2001a Dialogues for negotiation: agent varieties and dialogue sequences. In Meyer, J.-J. & Tambe, M. (eds.), *Intelligent Agent Series VIII: Proceedings of the 8th International Workshop on Agent Theories, Architectures, and Languages (ATAL 2001) (Lecture Notes in Computer Science, Vol. 2333)*. Berlin: Springer, pp. 69–84.

Sadri, F., Toni, F. & Torroni, P. 2001b Logic agents, dialogues and negotiation: an abductive approach. In Stathis, K. & Schroeder, M. (eds.), *Proceedings of the AISB 2001 Symposium on Information Agents for E-Commerce*.

Sadri, F., Toni, F. & Torroni, P. 2002 Abductive logic programming architecture for negotiating agents. In *Proceedings of the 8th European Conference on Logics in Artificial Intelligence (JELIA-2002) (Lecture Notes in Computer Science, Vol. 2424)*. Berlin: Springer-Verlag, pp. 419–431.

Samuelson, L. 1998 *Evolutionary Games and Equilibrium Selection*. Cambridge, MA: MIT Press.

Sandholm, T. 2002a Algorithm for optimal winner determination in combinatorial auctions. *Artificial Intelligence* **135**(1–2), 1–54.

Sandholm, T. 2002b eMediator: a next generation electronic commerce server. *Computational Intelligence, Special issue on Agent Technology for Electronic Commerce* **18**(4), 656–676.

Sandholm, T. & Crites, R. 1995 Multiagent reinforcement learning in the iterated prisoner's dilemma. *Biosystems* (Special Issue on the Prisoner's Dilemma) **37**, 147–166.

Searle, J. 1969 *Speech Acts: An Essay in the Philosophy of Language*. New York: Cambridge University Press.

Sen, S., Biswas, A. & Debnath, S. 2002 Believing others: pros and cons. *Artificial Intelligence* **142**(2), 179–203.

Sierra, C., Faratin, P. & Jennings, N. R. 1997 A service-oriented negotiation model between autonomous agents. In *Proceedings of the 8th European Workshop on Modeling Autonomous Agents in a Multi-Agent World (MAAMAW-1997), Ronneby, Sweden*, pp. 17–35.

Sierra, C., Jennings, N. R., Noriega, P. & Parsons, S. 1998 A framework for argumentation-based negotiation. In Singh, M., Rao, A. & Wooldridge, M. (eds.), *Intelligent Agent IV: 4th International Workshop on Agent Theories, Architectures and Languages (ATAL-1997) (Lecture Notes in Artificial Intelligence, Vol. 1365)*. Berlin: Springer-Verlag, pp. 177–192.

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press

Singh, M. P. 2000 A social semantics of agent communication languages. In Dignum, F. & Greaves, M. (eds.), *Issues in Agent Communication (Lecture Notes in Computer Science, Vol. 1916)*. Berlin: Springer-Verlag, pp. 31–45.

Spence, A. M. 1974 *Market Signaling: Informational Transfer in Hiring and Related Screening Processes (Harvard Economic Studies)*. Harvard University Press, Cambridge, MA.

Sycara, K. 1985 Arguments of persuasion in labour mediation. In *Proceedings of the 9th International Joint Conference on Artificial Intelligence*, pp. 294–296.

Sycara, K. 1992 The PERSUADER. In Shapiro, D. (ed.), *The Encyclopedia of Artificial Intelligence*. New York: John Wiley & Sons.

Tohmé, F. 1997 Negotiation and defeasible reasons for choice. In *Proceedings of the Stanford Spring Symposium on Qualitative Preferences in Deliberation and Practical Reasoning*, pp. 95–102.

Torroni, P. 2002 A study on the termination of negotiation dialogues. In Castelfranchi, C. & Johnson, L. (eds.), *Proceedings of the 1st International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2002), New York, USA*. ACM Press, pp. 1223–1230.

Torroni, P. & Toni, F. 2001 Extending a logic based one-to-one negotiation framework to one-to-many negotiation. In Omicini, P. P. A. & Tolksdorf, R. (eds.), *2nd International Workshop on Engineering Societies in the Agents World II (ESAW 2001) Lecture Notes in Computer Science, Vol. 2203)*. Berlin: Springer-Verlag, pp. 105–118.

Traum, D. 1999 Speech acts for dialogue agents. In Wooldridge, M. & Rao, A. (eds.), *Foundations And Theories Of Rational Agents*. Dordrecht: Kluwer, pp. 169–201

Varian, H. R. 1995 Mechanism design for computerized agents. In *Proceedings of the USENIX Workshop on Electronic Commerce, New York*.

von Neuman J. & Morgenstern, O. 1944 *The Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, NJ.

von Stengel, B. 2002 Computing equilibria for two-person games. In Aumann, R. & Hart, S. (eds.), *Handbook of Game Theory*, Vol. 3. Amsterdam: Elsevier, pp. 1723–1759

Vreeswijk, G. 1997 Abstract argumentation systems. *Artificial Intelligence* **90**(1–2), 225–279.

Walton, D. N. & Krabbe, E. C. W. 1995 *Commitment in Dialogue: Basic Concepts of Interpersonal Reasoning*. Albany, New York: SUNY Press.

Wooldridge, M. & Parsons, S. 2000 Languages for negotiation. In Horn, W. (ed.), *Proceedings of the 14th European Conference on Artificial Intelligence (ECAI-2000)*. John Wiley & Sons, pp. 393–397.

Wooldridge, M. J. 1997 Agent-based software engineering. *IEE Proceedings on Software Engineering* **144**(1), 26–37.

Wooldridge, M. J. 2000 *Reasoning about Rational Agents*. MIT Press.

Wooldridge, M. J. 2002 *An Introduction to MultiAgent Systems*. Chichester: John Wiley & Sons.

Wurman, P. R. 1999 *Market Structure and Multidimensional Auction Design for Computational Economies*. PhD thesis, University of Michigan.

Wurman, P. R., Wellman, M. P. & Walsh, W. E. 2001 A parametrization of the auction design space. *Games and Economic Behavior* **35**(1–2), 304–338.

Yokoo, M. 1998 The distributed constraint satisfaction problem: formalization and algorithms. *IEEE Transaction on Knowledge and Data Engineering* **10**(5), 673–685.

Young, M. A. 2001 *Rational Games: A Philosophy of Business Negotiation from Practical Reason*. Westport, CT: Quorum Books.

Yu, B. & Singh, M. P. 2002 Distributed reputation management for electronic commerce. *Computational Intelligence* **18**(4), 535–549.

Zeng, D. & Sycara, K. 1997 Benefits of learning in negotiation. In *Proceedings of the 14th National Conference on Artificial Intelligence and 9th Innovative Applications of Artificial Intelligence Conference (AAAI-97/IAAI-97)*. Menlo Park, CA: AAAI Press, pp. 36–42

https://doi.org/10.1017/S0269888904000098 Published online by Cambridge University Press