Negotiating Flexible Agreements by Combining Distributive and Integrative Negotiation

Quoc Bao Vo, Lin Padgham and Lawrence Cavedon
School of Computer Science and Information Technology, RMIT University, GPO Box 2476V, Melbourne, VIC 3001, Australia
E-mail: {vqbao,linpa,lcavedon}@cs.rmit.edu.au

Abstract

This paper presents an approach to automated negotiation between agents which attempts to combine the advantages of a co-operative value adding approach, with the reality that negotiating agents are also competing. We use the concept of a trusted mediator to facilitate openness regarding what one values, without disadvantaging oneself by revealing sensitive information (such as a reserve price) to the other party. Social science and management literature deals with negotiation between people, and so can be both more complex, and less well defined than automated negotiation between software agents. We take inspiration from the social science literature and develop a computational framework to support negotiating software agents. The framework includes recognition that agents are self interested, and therefore will manipulate the system to their advantage if possible. We include mechanisms to discourage this kind of manipulation in the form of a transaction cost associated with making only small concessions, and a bias in dividing the pie which is the gain from trade which favours the agent who is most "honest" in making offers.

Keywords: automated negotiation, integrative and distributive bargaining, BATNA, multi-criteria decision making

1 Introduction

Many decisions in our everyday life are increasingly being made by computers in a relatively autonomous fashion. For instance, the routing of telephone calls and data packets in telecommunication networks has been controlled and decided autonomously by computers. Computers make decisions and control over how loads in electrical grids will be balanced at times of peak demand. Similarly research is being done on how computers can react to, and control, automotive and airplane traffic in real time. In

*Corresponding author: Quoc Bao Vo, School of Computer Science and Information Technology, RMIT University, GPO Box 2476V, Melbourne, VIC 3001, Australia. Tel: +61 3 9925 3781. Fax: +61 3 9662 1617.
many cases, these decisions are generated as concerted efforts between several machines. Moreover, the involved machines might represent different entities or organisations, possibly with conflicting interests. Thus, there is a real need for these systems to negotiate with each other to come to agreements.

Nonetheless, negotiation is a complex task that needs to manage a balance between competition and collaboration. Most research in automated negotiation to date has focused on the competitive aspects. However, work by Dispute Resolution theorists in the social sciences has also focused substantially on how to achieve negotiated agreements that are of a high value to all parties (e.g., see [6, 19, 37]). Integrative, or interest-based, negotiation has been widely recognized as the more successful approach to the negotiation problem. In this work, we consider how some of the aspects of this approach can be represented in a computational framework, to allow its use in automated negotiation.

One of the difficulties with this is that, although the constructive and positive approach of integrative negotiation to creating value for all participants is extremely useful, negotiation still does have a competitive element, and negotiators can generally be assumed to be self-interested. In particular, Lax and Sebenius [19] argue that most negotiation actually involves both integrative and distributive bargaining. Moreover, these two aspects of a negotiation are most of the time intertwined (see the Negotiator’s Dilemma discussed later in this paper), resulting in a challenging decision-making problem for the negotiators. Consequently, we must ensure in our computational system that our approach be constructive and creating value, not lead to opportunities for manipulation for self-interest, or at least that such opportunities be recognized and controlled.

Our approach, at a high level, is to introduce a trusted third party, the mediator, who is guaranteed not to profit in any way from the outcome of the negotiation. The presence of this mediator does not remove all opportunities for manipulation, but does allow us to provide greater control, and also allows parties to reveal information to the mediator whilst keeping it from the other party. For example, to honestly disclose one’s reserve price (or for the buyer, top price to pay) to the other party would not be viable, as they would almost certainly use this information to their advantage. However, to disclose this information to a trusted mediator is a lot more viable. From a technical point of view, our framework is based on multi-criteria decision-making theory.

The use of a mediator in our framework helps facilitate more efficient outcomes in automated negotiations, which have become increasingly popular in electronic commerce [8, 23]. Our framework also forms a basis for Negotiation Support Systems (NSS) that provide support to human decision-makers who engage in real-life negotiations such as out-of-court settlements (see, for instance, [3, 7, 21, 22, 31]).

The overall aim of our approach is to increase the likelihood of finding an agreement if the possibility exists, to favour agreements that are of high value to both participants, and to favour agreements that are "fair", in the sense that advantage is shared between the parties in some controlled manner. In the rest of this paper, we explore

---

1 Informally, distributive bargaining is a competitive negotiation strategy that is used to decide how to distribute a fixed resource, such as money. Essentially, it contrasts with integrative bargaining in which the parties are trying to make more of something.
these ideas in more detail. We first provide some background on the two main approaches to negotiation: the integrative and the distributive approaches. In Section 4 we outline our computational framework that combines integrative and distributive negotiation with the assistance of a mediator. We then explore some details of the mediated negotiation process, focusing on mechanisms to avoid manipulation by either agent. Finally, we discuss the advantages that we consider this approach provides, compared to other work, and outline work in progress to evaluate whether in fact the approach does lead on average to higher-value, fairer transactions, even when agents are attempting to manipulate the situation for their own benefit.

2 Background

2.1 The “Getting to Yes” (GTY) approach

An approach that has had tremendous influence on negotiation theory in the social sciences, is the approach by Fisher and Ury [6], which focusses both on an integrative approach to the overall negotiation process, and also on the evaluation process of the individuals involved. In this work we focus primarily on the former aspect, though we also note that if agents adopted an evaluation strategy based on the principle of comparison with the “Best Alternative to Negotiated Agreement”, this would likely also assist in better outcomes of negotiation processes.

Much of Fisher and Ury’s work concerns discussion of the people behind the negotiation, how to take account of individuals, and how to effectively use what amounts to interpersonal skills. However there are a number of aspects of their model which can effectively be incorporated into a computational model, although such a computational model, which can be realised in software, of necessity lacks some of the creativity and nuances of human-based negotiation.

Some key aspects of Fisher and Ury’s [6] approach are as follows:

1. Focus on interests, not on positions:
   They stress the importance of recognising interests, both one’s own and the other party’s, rather than getting stuck in conflicting positions. They make the point that not all interests are necessarily conflicting. It is important to be clear and specific about one’s interests, while remaining flexible about possible positions.

2. Look at options for mutual gain:
   There are usually some aspects of a negotiation that are not in conflict, where there are win-win options. It is important to identify these and make them part of the agreement. They acknowledge though that this can be complex, in that disclosing options for mutual gain can also provide information that jeopardises the bargaining position regarding aspects that are in conflict.

3. Use objective criteria where possible:
   Most negotiations take place in a space where there are some standards or objective information available. It is appropriate to use this as part of a negotiation, although it may not always be clear exactly which objective information to use.
They give an example of an insurance company negotiating with a claimant about the value of a car. Although there are various possibilities with regard to the objective information - blue book value, new price, replacement value, etc. - the negotiation does not take place in an information vacuum.

4. Individuals need a comparison point:
They discuss the value of a clear “bottom line” in ensuring that an individual, in the heat of a negotiation, does not make an agreement that is not really in his/her best interest. They point out that although this can be important to avoid bad agreements, it is inevitably too rigid as it does not allow for flexibility in considering options that arise. They propose the concept of “Best Alternative To Negotiated Agreement” (BATNA) as an important concept for the negotiating parties. In order to evaluate a proposal, they argue that a party needs clarity on what their best option is if no agreement is reached - i.e. if the negotiation fails. All proposals should be considered in relation to the individual’s BATNA. It does not make sense to allow a negotiation to fail if the proposal offered is better than the BATNA. Clarity regarding one’s BATNA - as well as consideration of the other’s BATNA - is very important in the negotiation process. They also make the point that if one’s BATNA is strong, then advantage is gained by disclosing it, whereas if it is weak one is disadvantaged by disclosure. They also make the point that if both parties have a strong BATNA, it is possible that it is better not to reach a negotiated agreement.

Some of the concepts of Fisher and Ury’s approach have been incorporated into the integrative approach to negotiation, specifically with respect to trying to identify interests and to look for possibilities of mutual gain. Furthermore, research also shows that negotiators with attractive BATNAs feel less dependent on their opponent and achieve higher personal outcomes than negotiators with no (or less attractive) BATNAs [25, 26]. In addition, high BATNAs stimulate parties to search for integrative agreements that combine all parties’ aspirations, simply because parties are unwilling to settle for less than their alternative [35]. Thus, BATNA appears to be a good measure to be used in predicting negotiation outcomes as well as in negotiation analysis.

2.2 Distributive and Integrative Negotiation
By viewing the negotiation problem as a decision-making problem, most approaches are categorised either as descriptive (or behavioural) studies or normative studies. For instance, Walton and McKersie’s [37] analysis of qualitative differences between integrative and distributive negotiation types provided a basis for further descriptive studies of negotiation processes. They eventually led to the formulation of a number of prescriptions regarding strategies and tactics in distributive and integrative negotiations, their informational requirements, and even possible solutions. On the other hand, game theory, as opposed to descriptive studies, provides formal and normative approaches to model bargaining. Within this camp, bargaining theories are further divided up into two areas. Strategic bargaining games provide a formal description of the bargaining procedure and its influence on the equilibrium outcomes. This approach, however,
is based on the assumption that players behave non-cooperatively. As has been discussed throughout the present paper, this is not always plausible and usually leads to non-optimal negotiation outcomes. Cooperative bargaining theory, on the other hand, characterises the desirable properties of bargaining solutions by a set of axioms. However, the process leading to the outcome is not described in the solution.

Approaches developed from Fisher and Ury’s seminal work are generally espoused as positive, and leading to win-win situations, rather than win-lose situations which are described as resulting from the more traditional distributive approach. Basically, distributive negotiation predicts that one party can only gain at the other party’s expense. The parties are competitive and they claim value. In contrast, integrative negotiation is based on the premise that solutions can be found, during and because of the process, which reconcile the parties’ interests. The key characteristics that distinguish integrative negotiations from distributive ones are: creation of value; focus on interests and not positions; openness and exchange of relevant information, and even learning; and problem restructuring. Within the literature in social sciences, by both conflict theorists and management scientists, a number of monographs have been dedicated to the integrative negotiation approach (e.g., see [5, 6, 28]).

Within the literature of automated negotiation, Guttman and Maes [8] explore the approaches to automated electronic trading and claim that the distributive approach typically used has several destructive features, such as (a) hiding important merchant value-added services from consumer consideration, and (b) setting up the involved parties as opposing players in a win-lose game. They discuss a number of techniques such as multi-attribute utility theory, distributed constraint satisfaction, and conjoint analysis, which can help lead to a more integrative approach.

Our position is that while integrative negotiation is very important in reaching good agreements, in fact distributive and integrative negotiation are both necessary parts. This is a view supported by conflict theorists such as Lax and Sebenius [19], who refer to the integrative aspect as "creating value" and the distributive aspect as "claiming value". Negotiators face a dilemma in deciding whether to pursue a cooperative or a competitive strategy at a particular time during a negotiation. They refer to this problem as the Negotiator’s Dilemma.

Most of the work on automated negotiation tends to be clearly distributive in nature, looking at strategies and infrastructures that have nice game theoretic properties in terms of good behaviour by participating agents. This can be easily seen through the most prevalent works within the literature of automated negotiation, ranging from early and influential work on automated negotiation by Rosenschein and Zlotkin [30, 39], to Kraus’ [15] survey paper on negotiation and cooperation in multi-agent environments, on to the more recent line of research in automated negotiation led by Wooldridge and Jennings [4, 11] and Sandholm [18, 32]. Rosenschein and Zlotkin’s work explores several problem domains for automated negotiation, namely the task-oriented domain, the state-oriented domain, and the worth-oriented domain. Several negotiation strategies were also introduced, based on the assumptions that agents are self-interested and look to maximise their own utilities and do not attempt to cooperate to search for joint improvements. Kraus also looks only into strategic models of negotiation, in particular to solve the problems of resource sharing, task distribution, and coalition formation. Wooldridge, Jennings, and Sandholm, and their respective groups formulate the auto-
mated negotiation problem based on the bargaining game as defined by game theorists. This setting is purely distributive negotiation. This trend towards distributive negotiation in the literature of automated negotiation is perhaps most visible in the following statement in Wooldridge’s [38] review of Kraus’s [16] monograph “Strategic Negotiation in Multiagent Environments”:

[What] marks out negotiation as studied in multi-agent systems is the idea that the processes (agents) involved are self-interested: they are attempting to get the best deal for themselves that they can, and what is best for one is not necessarily best for the other.

Kersten [14] has explored approaches in terms of both integrative and distributive methods and has analysed differences between them. He does suggest some modifications to formal negotiation models to make them more flexible and useful. However to our knowledge there is no work in automated agent based negotiation which attempts to integrate the distributive and integrative approaches in a principled manner. It is this that we attempt to do.

The Negotiator’s Dilemma, of conflict between sharing information in order to achieve good cooperative solutions, and hiding information in order to maximise one’s competitive edge, is addressed, at least partially, by a mediator who allows both information disclosure, and information hiding. The use of a mediator is explored by Lin and Chou [20] who propose an approach based on Fisher’s Single Negotiation Text (SNT) device. The mediator introduces a proposed settlement and then the negotiating parties criticise it. The mediator then comes up with a new proposal that addresses (at least some of) the criticisms. However the Better Half (BH) strategy developed by them appears to be over-simplistic, and it is difficult to see how a mediator could work out new proposals without an integrative negotiation phase paving the way for the process of distributive negotiation.

2.3 Basic definitions

In this section, we define the concepts and their formulations which are derived from seminal works in negotiation analysis and multiple-criteria decision making [13, 27, 34].

1. **Negotiator** (aka. agent) is a party who is involved in a negotiation of discussion. Negotiators are denoted by $i$, ($i = 1, \ldots, n$).

2. **Attribute** (aka. issue, dimension, criterion) is the topic of discussion in a negotiation, such as price, length of warranty, delivery time, etc. Attributes are denoted by $j$, ($j = 1, \ldots, m$).

3. **Attribute value** (aka. option) is one of the alternative values that an attribute can take, such as $\$2500$ or ‘3-year-warranty’. Values of attribute $j$ can be denoted by $x_j$.

4. **Outcome** (aka. decision alternative) is a specific combination of values selected for one or more attributes (e.g., $[price = \$2500; warranty = ‘3-year-warranty’]$). Outcomes are denoted by $o = [x_j] \in \mathcal{O}$. 
5. **Interest** (aka. *objective*) is an attribute, or combination of attributes, that is of particular interest to the negotiator and allows to evaluate offers.

6. **Preference** indicates the importance of an interest in comparison with another interest.

7. **Trade-off** is an exchange process in which a negotiator gives up partly on the achievement of one issue so as to gain on another.

8. **Utility** is a measure based on the negotiator’s value function and some negotiation-specific regulations, such as participation fees, transaction costs, etc.

9. **Gains from trade** are the positive value attainable from the negotiated agreement. For instance, in a bargaining between a seller and a buyer over the price of an item, if the buyer values the item at $100 whilst the seller would accept as little as $70 for the item, then the gains from trade (also known as the *pie*, or the *cake* in the negotiation literature) in this negotiation is $30.

3 **Potential Advantages and Issues: An Example**

We introduce here a small example in order to illustrate aspects of integrative and distributive negotiation, and their benefits and limitations. We will continue to use this example as we describe our approach to negotiation.

The example we will take is one of a buyer and a seller negotiating over the purchase of a laptop. Using integrative negotiation to explore attributes of value to the various parties, we may be able to discover that in addition to the obvious attribute of price, there is a warranty attribute, where a long warranty is valued by the buyer, and is of little cost to the seller. Reciprocally, the payment type “cash” is perhaps valued by the seller, and is of no cost to the buyer. Consequently, with any given price agreement, a more efficient, or higher value, agreement can be reached by including a long warranty, and cash payment in the terms of sale. These added value options can possibly be found only by an exploratory phase to establish potential possible win-win aspects of the deal.

However, even if there are substantial win-win aspects that can be discovered and incorporated into an agreement, in most cases there is still some core aspect where the parties have conflicting interests. These aspects then still require some standard, or distributive negotiation, where some balancing between the competing interests is achieved. In our example (and virtually all trading examples), price is an attribute where it is virtually guaranteed that the buyer and seller will have conflicting interests:

---

2The concept of gains from trade is of course borrowed from game theory. In game theory, gains from trade appear in at least two ways. First, if a game has a positive value, then there are potential gains for the players. Second, in positive sum games, players may be able to coordinate their behaviour to produce gains for all. Both ways fit well into the discussions in this paper with the former corresponding to the fixed resource (i.e., the pie) to be shared between the negotiators during distributive negotiation. On the other hand, the latter corresponds to the negotiating parties trying to enlarge the pie by creating more value, i.e., integrative negotiation.
all else being equal, the buyer wants a low selling price, while the seller wants a high one.

There may occasionally be integrative solutions where there is in fact no conflict. For example consider two children negotiating over who will get a single orange. On exploration of what is valued, it turns out that one child wants to eat the orange, while the other wants the peel for baking a cake. A win-win solution is of course to share the orange in such a way that each child’s desires are fully met, as once the options are fully understood, there is no conflict. However most negotiation situations do have some element of conflicting desires.

If in our laptop sale example a distributive approach is taken from the start (even using multiple attributes), it is quite likely that a deal would be reached that did not necessarily contain the optimal values regarding payment type and warranty length. In some cases, if potential overlap is small, discovering and exploiting the non-conflicting attributes can make the difference between reaching an agreement or not. If for example the lowest viable price of the seller (assuming payment by credit card) is $2500, and the highest possible price for the buyer is $2400, then there would appear to be no possible agreement. However, if the use of credit card costs the seller 5%, then recognition of the fact that the buyer is willing to pay cash (and therefore save the seller $125) opens up the possibility of a deal.

Clearly integrative negotiation is valuable - but not enough in itself. However, naively combining integrative and distributive approaches also leads to problems. A buyer would prefer to keep the information that he is willing to pay cash, to obtain a final advantage once other negotiating strategies have been fully exploited. Disclosing this from the start potentially weakens his position. Maximising the use of the integrative approach requires openness and honesty. However the bargaining that is central to distributive negotiation relies on not disclosing all information. Thus there is an inherent conflict between the two approaches.

4 An Overview of the Proposed Approach

The approach that we propose recognises the view of conflict theorists that successful negotiation involves both creating value and claiming value. We propose two clearly separated phases – an integrative phase that explores and expands the value space, and a distributive phase that then determines the distribution details of that part of the value space where the parties have conflicting interests.

In order to address the conflict between sharing information and hiding information, we introduce a mediator with whom information is shared, thus addressing at least to some extent the issue of disclosure of information disadvantaging the disclosing party. However, it is still possible for the negotiating parties to be dishonest and/or manipulative in the information that they disclose, in order to gain a potential advantage. However, as we discuss below, such a strategy would risk derailing the negotiation to the detriment of all, including the “dishonest” agent. The mediator also controls the way the final pie of agreement is distributed: this is done in a manner that rewards “fairness”.

We aim, then, to define a model that results in:
• Efficient negotiations; if an agreement is possible, it will in fact be reached.
• Outcomes that are intuitively “fair”; gains from trade are distributed amongst the participants in an equitable manner.

In particular, the model is designed to penalize agents that attempt to manipulate the negotiation process to their unfair advantage.

Our model departs from previous approaches in three main ways:

1. We introduce a mediator into the negotiation process. The mediator is a trusted third-party agent that has no interest in the specific outcomes of the negotiation, in that it does not benefit according to the utility of either agent. A much-discussed example of the use of a mediator in a real world negotiation was the Camp David negotiation in September 1978 in which the negotiation between Egypt and Israel was mediated by the United States (see [27] for a detailed discussion).

2. We separate the integrative and distributive phases of the negotiation process. As described earlier, these two phases are often intertwined. One of the roles of the mediator is to manage each of these processes separately.

3. We use the mediator to reward “honest” negotiation behaviour (based on discrepancy between the agent’s reported reserve value and the offers she makes to reach an agreement) by biasing the final distribution of the pie on a deal being reached. We also apply a penalty based on the number of concessions made by an agent, in order to provide a disincentive from making small incremental concessions to their offers.

4.1 The Mediator

The Negotiation Mediator, or simply mediator, is an independent third-party agent that mediates information exchange and acts as a “referee” in the negotiation process. A critical property of the mediator is that it hold no interest from the specific outcome of the negotiation that it mediates; it may be compensated for conducting the mediation process, but the reward must not be based on any particular outcome.

The mediator is designed to both increase efficiency and help ensure fairness of the negotiation. It does so by the following means:

1. Acting as a decision-maker for integrative bargaining. Agents communicate their attributes of interest to the mediator, rather than to each other. The mediator calculates the intersection of attributes of value to both agents, effectively defining the scope of attributes over which distributive negotiation takes place. (This is discussed in further detail below.)

2. Acting as a conduit for communication. Full disclosure of an agent’s interests puts it at a disadvantage in a negotiation. Instead, agents communicate their

---

3Of course, for this case, it could be debated as to what extent the mediator had vested interests in specific outcomes.
interests and incremental offers/concessions to the mediator. The mediator com-
municates both the results of the integrative bargaining stage (see below), as
well as indicating at the end of each round of the distributive phase whether an
agreement/deal has been reached.

3. Acting as a referee, imposing penalties for “dishonest” bidding practices. Agents
are self-interested in advertising “reserve” prices that make the available size of
the pie as small as possible, i.e., leaving as little scope for negotiation as possible
while still allowing the possibility of a deal. Prior to the start of the distributive
phase of the negotiation, participants advertise to the mediator “reserve” prices
on each attribute of value.\textsuperscript{4} When a deal is reached, the mediator distributes the
pie so as bias against participants that have been “dishonest” by making offers a
long way from their nominated reserve.

4.2 Integrative Bargaining

Multi-attribute negotiation \cite{27} (also known as multi-issue negotiation) extends stan-
dard negotiation strategies by allowing participants to negotiate over multiple items
simultaneously, trading off concessions on different items which may have different
utilities for different participants. This has important benefits: the space of outcomes
is broadened, increasing the possibility of reaching a deal that is satisfactory to all
parties. The Camp David negotiation (in 1978) between Egypt and Israel \cite{27} is one
real-world example of a multi-attribute negotiation.\textsuperscript{5}

Given the complexity of having to reach agreements on multiple issues, simultane-
ous negotiation over all issues may not be always tractable. For instance, in the pres-
ence of incomplete information, non-cooperative bargaining theory cannot provide
any tractable solution to the problem of simultaneous (multi-attribute) negotiation. As
a consequence, game theorists mostly address the challenge by decomposing the prob-
lem into issue-by-issue negotiations. In addition to the problem of having to agree
on the negotiation agenda (on how to proceed with the issue-by-issue negotiation),
trading-off between different issues also becomes problematic. That is, all parties must
agree on one party’s concessions on an issue to allow this party to claim more shares on
future issues. For a survey on the literature on multi-attribute negotiations, the reader
is referred to \cite{17}.

Integrative bargaining is effectively the process by which the different attributes are
introduced to the negotiation. Standard approaches to mixed integrative- and distributive-
negotiation interleaves the two processes. However, this can prohibit reaching the op-
timally fair outcome for all parties: e.g., one agent may withhold information regarding
which attributes are important to it for the purpose of obtaining a better deal for itself.

Separating integrative and distributive negotiation into distinct phases removes the
possibility of this sort of manipulation. Each party is required to reveal to the mediator
which attributes are of interest to it at the start of the process, along with its proposed

\textsuperscript{4}Note that an agent may be dishonest about the reserve it advertises, but would then run the risk of the
mediator concluding that no deal was possible.

\textsuperscript{5}Other examples can be found in the literature: e.g., \cite{37}\cite{29}\cite{9}.
utility for each attribute. The mediator then defines the scope of the distributive bar-
gaining phase by deciding on the attributes over which negotiation will take place, 
without disclosing the proposed value of these attributes to the other parties. In this 
way, agents are unable to manipulate the scope of the negotiation, but neither are they 
disadvantaged by having their utilities revealed to other parties.

In summary:

- the mediator decides the scope of the negotiation;
- the mediator does not have all information: agents may not have been honest 
  with respect to stated value of the items under negotiation;
- participants don’t know anything about each other’s interests and perceived val-
  ues, other than what the mediator tells them the items of conflicting value are.

4.3 Distributive Bargaining

The distributive bargaining phase of the negotiation is the most usually studied in the 
literature on automated negotiation. Standard approaches to automated negotiation 
are based on game theoretic models under uncertain information. As discussed ear-
lier, in such approaches agents are considered to be self-interested and behave non-
cooperatively. Furthermore, in the presence of incomplete information, strategic sig-
nalling devices are introduced to facilitate agreements. The most common strategic 
devices used within the bargaining literature and a large part of automated negotiation 
literature are the use of time with discount factors, and the use of deadlines. For a spe-
cific information structure defined in a bargaining game, under some strict assumptions 
over the rationality of the agents as well as their unboundedly computational capability, 
research within this branch mainly focuses on the equilibrium analysis of the negotia-
tion game.

The important difference in our approach is the introduction of the mediator. The 
distributive phase begins with agents nominating their reserve price on each of the 
attributes under negotiation. While there is no way to guarantee or enforce that agents 
are honest with respect to the reserve they disclose, an agent that nominates a dishonest 
reserve (e.g., a buyer that discloses a reserve well below what she is really willing to 
pay) is in danger of sabotaging the possibility of any deal at all—the mediator may 
indicate that no deal is possible because there is no overlap between the nominated 
reserves. Since agents are self-interested in striking a deal (for somewhere inside their 
actual (unadvertised) reserve), then this would be a loss to that agent.

For our purposes here, we assume that in the case where there is no overlap between 
the reserve prices, the mediator cancels the negotiation altogether. In reality, of course, 
an agent may instigate a new negotiation process, nominating a reserve closer to her 
actual reserve; in this case, the mediator would impose a (heavy) penalty for this clear 
attempt to manipulate the process. However, for simplicity our current model only 
considers the case in which there is overlap between the reserve prices.

Note that all reserve information is kept from being made public. In particular, the 
reserves themselves are retained by the mediator but not disclosed to other agents. The
only uses of the reserve are (i) for the mediator to determine that a successful outcome is possible, and (ii) to estimate a “fair” distribution of the pie.

Further, actual offers are also not disclosed publicly, so as to reduce the possibility of manipulation. Rather than proposals and counter-proposals being presented directly to other participating agents, each agent presents its offer only to the mediator. Making an offer is multi-attributive and simultaneous, in that all agents make offers at the same time on all attributes under consideration. Recall that this may not include all attributes in the space: the mediator will have already determined the attributes of “value” during the integrative phase of the negotiation.

After each round of making an offer, the mediator announces whether or not a deal has been reached and, if so, determines the distribution of the pie. Details of each agent’s offers are not advertised, however, ensuring that agents are protected from having such information manipulated by others.

If a deal has not been reached in a given round, then each agent submits a new offer, which may involve a concession—i.e., an alteration of its previous offer that results in a reduced potential gain (e.g., a buyer offering a higher price or a seller willing to accept a lower price). Of course, an agent may have previously made an offer at its reserve price and be unable to concede any further. If two consecutive turns involve no concession by any party, then the mediator assumes that the possibility of a deal has passed and the negotiation is terminated. We describe below a penalty scheme that mitigates a potential attempt by an agent to manipulate this aspect of the protocol by continuing to concede (and thereby preventing the negotiation from terminating early) by offering very small concessions on its offers.

If the mediator ever receives offers that overlap, then the negotiation process is successfully terminated. The pie is that amount between the two final offers—e.g., between the price the buyer is willing to pay and the price the seller is willing to accept. The mediator informs the participants of the successful outcome, and fixes a final price that distributes the pie. The process for determining this price is discussed in the following section.

4.4 Promoting Efficiency and Fairness

One role of the mediator is designed to protect agents from being exploited by not publicly disclosing agents’ offers, and by making decisions, such as which attributes are of value to all, without disclosing the associated values. The protocol described in the previous two sections described the implementation of these.

At the end of a successful negotiation process—i.e., when offers from the agents have overlapped—the mediator (who is the only participant that knows all agents’ offers)—decides on the distribution of the pie. This distribution is calculated so as to bias against agents that have made offers far from their reserve: agents making offers closer to their reserve receive a corresponding larger portion of the pie;

The mediator also enforces a penalty designed to promote efficiency and fairness, and designed to prevent agents from manipulating the negotiation process. While each

---

6This situation is possible even if agents are not at their advertised reserve; e.g., if an agent concedes transaction costs (see Section 4.4) so that its effective reserve is altered by that amount.
agent has no direct information related to the value to which other agents hold the attributes under negotiation, an agent could still try to win a highly advantageous deal for itself by proposing offers which are a long way from its reserve, and by then making extremely small incremental concessions on its offer position. For example, a buyer may increase its offer on an item by the smallest amount possible at each round (e.g., making new offers on a house at one cent increments!).

We propose a penalty scheme to reduce the advantage of such behaviour:

- To disincentivise agents from making tiny incremental concessions, a transaction penalty, or concession penalty, is imposed for each round in which an agent concedes.

The transaction penalty is crucial not just for ensuring greater efficiency of the negotiation process, but also to help ensure fairness. The pie-distribution protocol has less effect if agents make only small concessions: if this were the case, then the size of the pie would be correspondingly small and a “dishonest” agent would have little to lose from the biased distribution of the pie. However, the transaction cost encourages all agents to reach an agreement early, making it advantageous to increase their offers, while ensuring less to lose if competing agents don’t do the same (due to the way the pie is distributed).

Of course, if an agent starts by providing a highly dishonest reserve price (i.e., by nominating a reserve that is not close to their actual reserve value) then the loss in the pie-distribution can be minimised. Similarly, an agent may simply not concede (even though it hasn’t reached its reserve price) and not incur a transaction cost. However, in both these situations the agent would risk having the mediator terminate the negotiation (either because there was no possible solution to start with, or if the other agent also happened to not concede). Since our assumption is that it is in the agents’ interests to reach a deal (otherwise they wouldn’t be participating in the negotiation to start with), these are both risky strategies.

The degree to which fairness and efficiency are impacted by these penalty schemes is dependent on the actual penalty values. This is discussed further below.

5 The Negotiation Process Using the Mediator

In this section, we set out to describe a negotiation procedure that allows the negotiators to go through both phases of a negotiation, namely value creation and value claiming. Essentially, a negotiation procedure characterises a plan of interaction between the parties confronted with a negotiation problem. We make several assumptions to make applicable several techniques from the Operations Research literature for searching for joint gains. We assume that \( O \subseteq \mathbb{R}^m \): i.e., the attributes take values from the set of real numbers \( \mathbb{R} \). We further assume that the set \( O \) is convex. We also formalise the negotiating parties’ interests using the objective vector function, for each agent \( i \), \( \hat{f}_i = [f_{ij}] : O \rightarrow Y_i, \ (j = 1, \ldots, q_i) \), where \( q_i \) is the number of objectives of agent \( i \), and \( Y_i \subseteq \mathbb{R}^{q_i} \) limits the feasible ranges for the objective functions \( f_{ij} \) of agent \( i \). A utility function \( u_i : Y_i \rightarrow U_i \), where \( U_i \subseteq \mathbb{R} \), can be defined for each agent \( i \). We also assume that the functions \( \hat{f}_i \) and \( u_i \) are concave.
**Remark:** We are deviating here from the standard representation in which the utility functions map from the set of outcomes to $\mathbb{R}$. By using the objective functions, we are able to separate the agents’ interests from the attributes of the outcomes. Attributes are often, especially in auctions, considered equivalent to objectives. Note, however, that negotiators may introduce and discuss spurious attributes in order to achieve preferable outcomes on other attributes. In many negotiations interests are not revealed and are defined over several attributes. As discussed thoroughly by Keeney and Raiffa [12], it is quite likely that an agent’s multiple objectives will conflict with each other in that the improved achievement with one objective can only be accomplished at the expense of another. For instance, most businesses and public services have objectives such as “minimise cost” and “maximise the quality of services.” Since better services can often only be attained for a price, these objectives conflict.

**Example 1** We reconsider the example discussed in Section 3. In this example, we have the following three attributes comprising the decision alternatives (i.e., outcomes) in this negotiation: $p$ for price, $l$ for the length of warranty (assuming that the value space for this attribute is the interval $[0, 5] \subseteq \mathbb{R}$), and $a$ for the amount paid by credit card. Note that we have deliberately reformulated these attributes to allow them to take values from the set of real numbers $\mathbb{R}$. That is, instead of using a discrete decision variable to represent the payment method (cash vs. credit card), we have translated it into a continuous variable $a$.

In this example, the seller $S$ has two objectives: “minimise cost,” $f_{S,c}$, and “maximise revenue,” $f_{S,r}$. We define: $f_{S,r} = (p-a)+0.95 \times a$, and $f_{S,c} = c+30 \times l$, where $c$ is the seller’s cost of providing the laptop and $30$ indicates the seller’s cost of providing a warranty for the laptop per year. On the other hand, the buyer $B$ has also two objectives: “minimise cost,” $f_{B,c}$, and “optimise after-sale services,” $f_{B,s}$. We define: $f_{B,c} = p$, and $f_{B,s} = 100 \times l$. Thus, the buyer actually values each year of warranty at $100$, and it is indifferent to him whether any amount would be paid by credit card. Now, assume that $u_B(f_{S,c}, f_{S,r}) = f_{S,r} - f_{S,c}$, and $u_B(f_{B,c}, f_{B,s}) = v + f_{B,s} - f_{B,c}$, where $v$ is the buyer $B$'s valuation of the laptop’s worth when it comes with no warranty. Consider the case of $v = c$. That is, the buyer’s valuation of the laptop is the same as the seller’s cost to provide it. When $l = 0$ and $a > 0$, then no agreement can be reached as it costs the seller more to provide the laptop than it is worth to the buyer. However, when $l = 5$, a maximum gain from trade of $350$ (provided that $a = 0$) can be generated. That is, the buyer values the laptop with a 5-year warranty at $(c + $500) while it only costs the seller $(c + $150) to provide this deal. Thus, any price $p$ between $(c + $150) and $(c + $500) would give both players positive utilities.

We now discuss several mechanisms to help the negotiating parties reach such efficient outcomes.

**5.1 Information Structure of Negotiation**

Research has shown that cooperation in general brings better pay-offs to all parties in terms of mutual benefits or by attaining common goals and thus fully realising potential gains (see [1, 2, 10, 24, 33]). However, it is still the case that in real world situations
people do not always behave cooperatively, especially during negotiations, even when the goal is to search for joint gains. This is evidenced by the negotiators’ frequent failure to achieve efficient agreements in practice [27, 34] and their unwillingness to disclose private information due to strategic reasons. While further empirical and systematic studies need to be performed to explain this behaviour, it is also clear why a party would behave non-cooperatively when she enjoys distinct unfair advantages against her opponents. For instance, the advantageous party could have complete information about her opponents’ private information such as their objectives and utility functions and all of their alternatives outside of the negotiation. If this is the case, the informed party is unlikely to cooperate with the mediator, but would prefer to take advantage of this additional information to either manipulate the outcome of the negotiation, or to claim the largest possible share of the pie to herself. If this happens to be the information structure of the negotiation, a mediator is unlikely to be able to play a constructive role in the negotiation since the advantageous agent will dictate the negotiation process by forcing other parties to make all possible concessions.

Fortunately, such an asymmetric information structure is rather rare in most real-world situations and in automated negotiation. Therefore, we assume in the remainder of this paper that the information structure of the negotiation is symmetric, giving no particular unfair advantage to any of the agents. Thus, all agents have the incentive to cooperate with the mediator to extract all possible gains attainable from the negotiation.

### 5.2 Using the Mediator at the Integrative Phase

The integrative phase of the negotiation involves deciding which attributes of value can be ignored during the distributive phase due to there being no conflict of value, and which values need to be negotiated over. There are a number of different strategies possible here: [36] discusses some of these strategies as well as providing formal details.

In this phase, agents provide information about their utility functions and objectives to the mediator, and information on some attributes as they deem appropriate. Using this information, the mediator determines the attributes for which an optimal value can be found, and those need to be resolved by further negotiation (i.e., at the distributive phase). This determination is partly decided by the amount of information that the agents choose to disclose.

For example, we refer to the laptop purchasing example above. The Seller and Buyer agents disclose their respective utility functions $u_S$ and $u_B$, along with their objective functions (i.e., “minimise cost” and “maximise revenue” for Seller; “minimise cost” and “maximise after-sales service” for Buyer). The Seller also discloses information about the cost of providing the warranty ($30 per year) and the cost incurred by a credit card transaction (5% of price paid). The Buyer discloses information about her valuation of the warranty ($100 per year) and indifference between paying cash or credit card. However, they do not disclose $c$, the cost to the Seller of providing the laptop, or $v$, the valuation by the Buyer of the laptop.

Given this information, the mediator is able to maximise the joint utilities for all parties (i.e., $u_S + u_B$), while holding $c$ and $v$ constant (since it has no information on $c$ and $v$), by choosing values for the length of the warranty $l$ to be 5 (the maximum
allowed) and fixing the payment to be cash rather than credit card. Note that the actual cost of providing the warranty is not fixed, but is incorporated into the negotiation over the price of the laptop: i.e., the distributive phase becomes a negotiation over the price of a laptop and 5-year-warranty bundle, with the payment being in cash.

Of course, the attributes and values that can be set by the mediator prior to the distributive phase of the negotiation depend on the information that the agents choose to disclose: if one of the agents had chosen not to disclose any information about the warranty, say, then the mediator would not have been able to fix the length of warranty at this stage, and this would have become part of the negotiation at the distributive phase.

As we described earlier, the advantage of agents disclosing information about the “non-core” attributes during the integrative phase is so that the distributive phase can be simplified, and focused on the specific attributes that matter (in our case, the price of the laptop/warranty bundle). In the extreme cases such as those discussed in Section 5.1, agents do not disclose enough information for the mediator to set any of the attributes, and the distributive phase degenerates into standard multi-attribute negotiation [27]. See [36] for details on alternative approaches to how the integrative phase may be managed.

At this point, the mediator and negotiating parties are almost ready to begin the distributive phase of the negotiation. The final step is for each agent to disclose a reserve to the mediator for each attribute that is the subject of the negotiation: in our case, it is the price of the laptop/5-year-warranty bundle. Of course, an agent may choose to disclose a dishonest reserve in order to try to obtain an advantage: e.g., the Seller may nominate a reserve that is higher than the price she is really willing to accept; conversely, the Buyer may specify a reserve that is lower than the price she is really willing to pay. Potential consequences for false disclosures are discussed below. However, as long as there is a gap between these reserves (i.e., the Seller’s reserve is lower than the Buyer’s reserve) then the mediator will open the distributive phase by inviting opening offers.

5.3 Using the Mediator at the Distributive Phase

The previous section discussed the integrative process, whereby the mediator determined which attributes were in conflict and which attributes could be fixed and solved independently. For example, fixing a 5-year warranty and cash-payment was deemed by the mediator to be a good outcome for both parties.

The conflicting attribute—in this case, (cash) price $p$ of the laptop plus warranty—is the attribute on which the Seller and Buyer make proposals. In general, there may still be multiple attributes to negotiate over at the distributive phase; for simplicity we illustrate the process with a single attribute.

5.3.1 Reserve Disclosure.

As part of the integrative phase, the two parties released information to the mediator relating to their reserve price. Let the Seller’s disclosed reserve price for the laptop-

---

7Further details on the optimisation process can be found in [36].
plus-warranty bundle be $p_s$, where $p_s$ is likely to be greater than $c + 150$ (where $c$ is the cost of the laptop to the Seller and $150$ is the cost of offering the 5-year plan).\(^8\)

Similarly, we denote the Buyer’s disclosed reserve price for the bundle as $p_B$, where $p_B$ is likely to be below $v + 500$ (where $v$ is the Buyer’s valuation of the laptop and $500$ is her valuation of the 5-year warranty).

Of course, either or both agent may have been dishonest in the reserves they disclosed to the mediator: the Seller’s actual reserve for the bundle is $p_s^a \leq p_s$, while the Buyer’s actual reserve is $p_B^a \leq p_B$. Under an ideal honest disclosure a deal would be possible so long as it is the case that $p_s^a \leq p_B^a$. However, the mediator will call off the negotiation unless $p_s \leq p_B$, which is potentially a more narrow condition. This is the risk with disclosing a dishonest reserve price to the mediator: the mediator may call off the negotiation, to the detriment of both agents, in a situation where a deal may actually have been possible. For example, suppose the Seller would potentially accept any price above $2600$, covering costs of laptop and warranty and providing a (minimal) profit, while the Buyer is willing to pay up to $2900$ for laptop plus warranty. In this case a deal is clearly possible. However, if the Seller informs the mediator that she requires a price of $3000$ at minimum, then the mediator terminates the negotiation and the opportunity of a deal is missed. Similarly, if the Seller tells the mediator she insists on $2750$ and the Buyer also informs the mediator of a maximum price of $2700$, then once again the mediator terminates the process.

5.3.2 Making Offers and Concessions.

Unlike in standard distributive negotiation, each agent’s offer is not publicly announced but only disclosed to the mediator. At each step, the mediator announces whether agreement has been reached; if not, new offers are invited, with each agent choosing either to re-offer the same amount or choosing to make a concession (i.e. make an offer that is closer to its reserve).

If an agent chooses not to make a concession, then it runs the risk of having the negotiation terminated. That is, after a round in which no agreement has been reached and no concessions have been made by any of the agents, the mediator announces that a possible deadlock could be reached and invites the negotiating parties to make concessions to avoid the negotiation from being called off. The mediator then calls a halt if no party makes any further concession in the subsequent round.

On the other hand, making a concession involves paying a penalty, as described above. This effectively lowers the value of the pie for the penalised party. Equivalently, the mediator can view this as the reserve effectively being adjusted (higher for a Seller, lower for a Buyer) by the amount of the penalty.\(^9\) For example, consider again our example involving the laptop and warranty, whereby the Seller nominates a reserve price of $2600$ and the Buyer a reserve of $2900$, and suppose the policy in place charges $10$ for each concession. After a concession from each side, the mediator will believe that the Buyer would now only be willing to pay up to $2890$, and that the Seller

\(^8\)Of course, in some situations sellers may actually drop selling price below cost.

\(^9\)Of course, an agent may have disclosed a dishonest reserve, but for the purposes of deciding when agreement is no longer possible it amounts to the same thing.
would require at least $2610 to cover costs and to protect profit margin. If each party has conceded, say, 10 times, then the penalties imposed leave little margin for trading.

In general, suppose the Seller nominates a reserve $p_s$ and has offered a concession on $n$ occasions, and suppose the Buyer has nominated a reserve of $p_B$ and offered $m$ concessions. If the penalty for each concession is $c$ then the mediator only believes a deal is possible while $p_s + n \cdot c \leq p_s - m \cdot c$. Again, the mediator’s information about actual reserve may be incorrect, but this is the information the mediator acts on.

The precise value of the concession penalty is varied across markets, and may potentially be fixed by the mediator at the start of the distributive negotiating phase. Clearly, the higher the value of this penalty, the greater the incentive to reach agreement quickly, and the higher the disincentive for agents to attempt to win a larger share of the pie by small incremental concessions. In our scenario, a concession penalty of $100 would highly motivate both agents to open with an offer close to their reserve! (Opening offers incur no penalty, of course.)

If one agent does offer close to the reserve and the other “bluffs” in the hope of winning the larger share of the pie, then the first agent may also be forced not to concede further while the other agent will be forced to do so and incur the penalty. For example, supposing the exhorbitant penalty of $100, if the Buyer offered $2850 (close to its reserve of $2900) while the Seller offered $2900, then the Buyer would be forced not to change its offer (or incur a penalty that took it past its stated reserve) while the Seller would expect to concede (and thereby incur the penalty). If the Seller next offered, say, $2800, the mediator would indicate that an agreement had been reached and proceed with distributing the pie of value $50 (the difference between the final offers). The Seller would not only receive $100 less than the final price determined by the mediator, but would also be disadvantaged in the pie-distribution since it made offers further away from its stated reserve: this issue is discussed below.

It should be noted that, in general, there is the potential for a variant of the Prisoner’s Dilemma here: one agent may take the risk of incurring the concession penalty and “bluff”, in the expectation that the other will make an opening offer close to its offer, whereas a joint-optimal strategy is for both agents to make “fair” offers and not concede any penalty. Setting a higher penalty lowers the potential success of such a strategy; however, setting the penalty too high may result in too much value being drawn from the negotiation. We are currently performing experiments to determine optimal penalty-setting strategies for different negotiation scenarios.

### 5.3.3 Distributing the Pie.

Since the agents’ offers are not publicised to each other, an agreement may be reached with significant overlap between the respective offers. This overlap, or pie, needs to be distributed. The distribution process is designed to promote fairness by biasing towards agents that have been making offers more “honestly”, in the sense of being closer to their nominated reserve. This could be done in a number of ways. Here, we assume that the mediator simply distributes the pie by proportioning the overlap between final offers according to how far each is from the corresponding agent’s nominated reserve price. A possible alternative is to take agents’ history of offers over the negotiation into account.
Following on with our example, suppose that the mediator detects a deal when the Buyer offers $2850, with the Seller having conceded to $2800, leaving a pie of $50. This leaves the Buyer $50 from its nominated reserve of $2900 and the Seller $200 from its reserve of $2600. The mediator therefore sets the final price at $2810, i.e., so that 20% of the pie is allocated to the Seller and 80% to the Buyer.\(^{10}\)

While this strategy can minimize advantages gained by aggressive and/or dishonest negotiation strategies, it can never remove them of course, since the final price will always be further from its reserve than an agent’s final offer. However, this is only reasonable: an agent that negotiates successfully reaps the rewards of that success; but some of the advantages of a “dishonest” strategy of making offers a long distance from its reserve are dampened.

An important final point is that the protocol for distributing the pie does not take any concession penalties into account: i.e., the reserve prices as originally advertised by the agents to the mediator are used for calculating proportional distribution. Since agents are aware of all concession penalties they have incurred, we can assume that the final price set by the mediator will always be one that the agent is prepared to accept. In particular, assuming the Seller of our example conceded once, then her final remuneration for the laptop and warranty is $2710, i.e., the mediator-set final price of $2810 minus the concession penalty of $100. Assuming the Buyer did not concede, then she pays the set price of $2810, and the penalty of $100 is paid to the marketplace.

6 Discussion and Conclusion

In this paper we have described a framework which recognises that negotiation has a greater chance of success if two different approaches to negotiation are combined. The traditional distributive approach of trying to reach agreement while coming from opposite directions, is complemented by an integrative approach which tries to discover all possible aspects of the situation which could be helpful in reaching an agreement.

The integrative approach is well developed and accepted by conflict theorists and management scientists. We have attempted to take some of these ideas and incorporate them into a computational approach suitable for automated negotiation. A number of the key points suggested in the seminal work of Fisher and Ury [6] have been incorporated into our framework. The importance of focusing on interests, not positions is addressed in our framework by having an initial phase where agents openly reveal their interests to the mediator. Employing a mediator avoids the potentially negative effects of disclosing information to the person you are negotiating with. The key point of looking at options for mutual gain is achieved by having the mediator attempt to find optimal mutual gain solutions for as many attributes as possible of those disclosed by the agents, prior to entering the distributive phase of the negotiation. Fisher and Ury also suggest that objective criteria should be used where possible. While we have not explicitly discussed this in our framework, it would certainly be possible for the mediator to incorporate information regarding objective criteria within a particular domain, into the final stage of determining how the available pie should be distributed.

\(^{10}\)I.e., $50:$200 is 20:80.
once an agreement is reached. In fact knowledge of objective criteria could be used by the mediator to continue pushing the agents for concessions, even after an agreement was possible, in the interest of reaching an agreement that was fairer according to some objective criteria.

The final important point made by Fisher and Ury, regarding the importance of the agents being clear about their Best Alternative to Negotiated Agreement, or BATNA, has to do with the individual agents’ utility function, and is therefore not covered in this paper. However, if agents do have a clear and well defined BATNA, and reveal this to the mediator in the initial phase, then our framework makes it possible for the mediator to determine immediately that a negotiated agreement is actually not in the interests of the negotiating parties.

While the framework that we have described does attempt to control manipulation using small concessions, as well as rewarding agents that propose solutions closer to their real reserve, there is still room for manipulation. Unlike some of the game theoretic distributive approaches, we have not defined a strategy which ensures that an agent is unable to realise gains by playing a manipulative strategy. However, we believe that across many transactions, the framework we have proposed does favour agents using open and non-manipulative strategies.

We have implemented the framework and are in the process of designing experiments to explore this issue, focussing on agents in different kinds of environments with regard to the characteristics of the other agents. It may well be that a manipulative agent operating in a world where all other agents are open and non-manipulative does realise an advantage, but once the environment is more heterogeneous this advantage may disappear, and in fact become a disadvantage in terms of numbers of successfully concluded negotiations, as well as profits. As we gain greater understanding of the nuances of the two phases, the basic framework can possibly be adjusted slightly to be more robust with respect to potential manipulation. However we believe the basic framework we have provided will prove to be both powerful and robust. It combines integrative and distributive negotiation in a well defined manner, managing at least to a large extent the anomaly that good integrative process requires openness and honesty, whereas in the competitive distributive process agents can be in a more powerful position if they do not disclose everything.

Acknowledgement

This work is supported by the Australian Research Council under grant number DP0663147. We are also grateful to the Australian Tourism Data Warehouse (ATDW) for access to their live tourism information for delivery by our software agents.

References


