Int. J. Human-Computer Studies (1998) **48**, pp. 125–141

Bayesian Learning in Negotiation*

Dajun Zeng Katia Sycara
Graduate School of Industrial Administration
The Robotics Institute
Carnegie Mellon University, Pittsburgh, PA 15213
Voice: (412) 268-8815 Fax: (412) 268-5569
Dajun.Zeng@cs.cmu.edu katia@cs.cmu.edu

October 13, 1998

Abstract

Negotiation has been extensively discussed in game-theoretic, economic, and management science literatures for decades. Recent growing interest in autonomous interacting software agents and their potential application in areas such as electronic commerce has given increased importance to automated negotiation. Evidence both from theoretical analysis and from observations of human interactions suggests that if decision makers can somehow take into consideration what other agents are thinking and furthermore learn during their interactions how other agents behave, their payoff might increase. In this paper, we propose a sequential decision making model of negotiation, called Bazaar. It provides an adaptive, multi-issue negotiation model capable of exhibiting a rich set of negotiation behaviors. Within the proposed negotiation framework, we model learning as a Bayesian belief update process. In this paper, we present both theoretical analysis and initial experimental results showing that learning is beneficial in the sequential negotiation model.

1 Introduction

Recent growing interest in autonomous interacting software agents and their potential application in areas such as electronic commerce (Sandholm & Lesser 1995) has given increased

* This research has been sponsored in part by ONR grant #N00014-95-1-1092, by ONR Grant No. N00014-96-1-1222, by ARPA Grant #F33615-93-1-1330, and by NSF grant #IRI-9612131.

importance to automated negotiation. Much DAI and game-theoretic research (J. & Zlotkin 1994, Osborne & Rubinstein 1994) deals with coordination and negotiation issues by giving pre-computed solutions to specific problems. There has been much research reported on developing theoretical models in which learning plays an eminent role, especially in the area of adaptive dynamics of games (e.g., (Jordan 1992, Kalai & Lehrer 1993)). However, to build autonomous agents that improve their negotiation competence based on learning from their interactions with other agents is still an emerging area.

We are interested in developing autonomous agents capable of reasoning based on experience and improving their negotiation behavior incrementally. Learning in negotiation is closely coupled with the issue of how to model the overall negotiation process, i.e., what negotiation protocols are adopted. Standard game-theoretic models (Osborne & Rubinstein 1994) tend to focus on *outcomes* of negotiation in contrast to the *negotiation process* itself. DAI research (J. & Zlotkin 1994) emphasizes special protocols articulating compromises while trying to minimize the potential interactions or communications of the involved agents. Since we are motivated by a different set of research issues, such as including effective learning mechanisms in the negotiation process, we adopt a different modeling framework, i.e., a sequential decision making paradigm (Bertsekas 1995, Cyert & DeGroot 1987).

The basic characteristics of a sequential decision making model are: (1) there is a sequence of decision making points (different stages) which are dependent on each other, and (2) the decision maker has a chance to update his knowledge after implementing the decision made at a certain stage and receiving feedback so that he can make a more informed decision at the next stage. The following observations support our choice of sequential decision making as the baseline negotiation model. First, most negotiation tasks involve multiple rounds of exchanging proposals and counter-proposals. A sequential decision making framework provides readily available constructs to model the iterative nature of inter-agent interactions. Second, negotiating agents indeed receive feedback after they offer a proposal or a counter-proposal in the form of replies from the recipient agent(s). Third, a sequential decision making framework supports an open-world approach. An agent does not need to have a complete world model at the outset of negotiation. Whenever new information comes in, irrespective of whether the agent learns the new knowledge by itself or some other information sources become available, it can make use of the newly acquired knowledge at the next decision making point. The agent can handle in the same manner more difficult situations where agents not only do not have complete information, but also the environment and other agents might be constantly changing. Last but not least, learning can take place naturally in a sequential decision making framework. This type of online incremental learning behavior is highly desirable in an automated negotiation program.

In this paper, we propose such a sequential decision making model, called Bazaar, which is able to learn. We address multi-agent learning issues in Bazaar by explicitly modeling beliefs about the negotiation environment and the participating agents under a probabilistic framework using a Bayesian learning representation and updating mechanism. We also report our initial experimental results in a simple bargaining scenario. Our ultimate research goal is to develop an adaptive negotiation model capable of exhibiting a rich set of negotiation behaviors with modest computational effort.

2 A Survey of Existing Negotiation Models

Traditional single-agent decision making models typically assume that the decision maker has complete knowledge of (1) his own preference ordering or utility function, and (2) the probabilities associated with the various outcomes. When multiple agents are involved, such as in negotiation, the introduction of *strategic* interaction, however, complicates this picture. In making his decision, the rational individual must take into account the probable choices of others, whose choices are in turn contingent upon his own. This leads to the well known *outguessing regress* (Young 1975) where no accurate prediction or confident expectation about the individual choices can be produced. Therefore, the central theme of all negotiation models is avoiding this *dilemma* involved in strategic interaction.

In order to circumvent the *outguessing regress* of strategic interactions, game-theoretic models make the following restrictive assumptions[1]: (1) Both the number of players and their identity are assumed to be fixed and known to everyone. (2) All the players are assumed to be fully rational, and each player knows that the others are rational (common knowledge). Each player's set of alternatives is fixed and known. (3) Each player's risk-taking attitude and expected-utility calculations are also fixed and known to each and every individual involved in decision making. These assumptions limit the applicability of game-theoretic frameworks for solving realistic problems. Another important limitation of game-theoretic models is that these models are fundamentally static in the sense that they primarily focus on *outcomes* in contrast to negotiation *processes*. The search for determinate rational decisions within

[1]It should be noted that some of the very recent game theoretic models are directly motivated by considerations of dropping or relaxing some of these assumptions. Although there has been interesting progress reported in the literature (e.g., (Jordan 1992)), the fundamental framework and methodology of game theory remains almost the same and it might be too early to tell whether these new results will reshape the current game-theoretic framework.

the framework of game theory has not led to a general model governing rational choice in interdependent situations. Instead, it has produced a number of special models applicable to specific types of interdependent decision making. For instance, the most celebrated solution concept, the von Neumann-Morgenstern solution, is based on the fact that in a two-person, zero-sum game an outguessing regress can be avoided by assuming (not unrealistically) that one player knows that his opponent will "do his worst", whatever strategy he himself selects. Analyses of the N-person cooperative game circumvent the difficulties associated with strategic interaction in a different way by introducing detailed decision rules concerning such things as the relative *power* of the players (e.g., the Shapley value and the Nash solution(Luce & Raiffa 1957, Nash 1950)).

Some game theorists (e.g., (Harsanyi & Selten 1972)) have sought to achieve determinate solutions for nonzero-sum games by introducing the notion that each player may be able to assign *subjective probabilities* to the choices of the other participant. In other words, it is possible to suppose that each individual proceeds in some subjective fashion to estimate the probable choices of the other player. In essence, the individual acquires information in the process so that his choice problem reduces to a situation that is fundamentally analogous to a game against nature as in a traditional single-agent decision making situation. We view this line of research as more closely coupled with sequential decision making view of negotiation rather than orthogonal game models.

To a large extent, these theoretical models are not concerned with computational issues, i.e., how to deal with inevitable practical complexities that do not have proper analytic representations and therefore haven't found their way into the models. Some of the AI models, in this sense, can be understood as bridges between applications and abstract theoretical models. Playing games (e.g., chess, go) has been one of the major foci of AI. For certain games, game theory is able to provide a theoretically sound mathematical solution and winning strategy. The existence of the solution, however, does not guarantee that the player can find the solution. AI models and programs help the players locate an approximate solution strategy according to bounded rationality principles by utilizing heuristic search, heuristic evaluation, and learning techniques(Russell & Wefald 1991, Rich & Knight 1991). Along with the emergence and development of DAI techniques, there has been increasing interest in using AI methodology and frameworks in negotiation modeling. (Sycara 1990) enriched the negotiation model by integrating AI planning, Case-Based Reasoning and other decision-theoretic techniques. Multi-agent resource allocation as a special case of negotiation has been extensively explored by (Kraus & Subrahmanian 1995), in which logic framework and time constraints are taken into consideration within the traditional framework of game theory. Some recent work (e.g., (Sen & Sekaran 1995, Sandholm & Lesser 1995)) in the context of Distributed AI addresses multiagent learning issues in various settings. Our work differs from others by explicitly modeling negotiation as a sequential decision making task and using Bayesian updating as the underlying learning mechanism.

3 Sequential Decision Making with Rational Learning

Our overall research goal is to develop a computational model of negotiation that can handle multi-agent learning and other complicated issues (e.g., multi-issue multi-criteria negotiation) that do not have straightforward and computationally efficient analytic models. We believe that a useful computational model of negotiation should exhibit the following characteristics: (1) The model should support a concise yet effective way to represent negotiation context. (2) The model should be prescriptive in nature. (3) The computational resources required for finding reasonable suggestions/solutions should be moderate, sometimes at the cost of compromising the rigor of the model and the optimality of solutions. (4) The model should provide means to model the dynamics of negotiation. (5) The model should also support the learning capability of participating agents.

Motivated by these desirable features, we have developed Bazaar, a sequential decision making negotiation model that is capable of learning. We describe how the proposed model works in a simple negotiation scenario for illustrative purposes before we present the formal description of Bazaar.

Suppose two computer programs are negotiating on behalf of their users in a supply chain management scenario. `Agent 1` is the producer (supplier)'s agent and `Agent 2` is the buyer's agent. These two agents are involved in a negotiation process where a detailed contract concerning product mix, delivery date, price, etc., is expected to be achieved. The overall negotiation process can be modeled as exchanging proposals and counter-proposals, as typically happens in human negotiations.

Let us first view the negotiation from the supplier, i.e., `Agent 1`'s point of view. We ignore the problem associated with locating potential buyers and assume that the existence of `Agent 2` is known to `Agent 1`. We also assume a communication channel between `Agent 1` and `Agent 2` is readily available. At the outset, `Agent 1` needs to come up with a solution package detailing its offer with respect to product, price, delivery date, quality, etc. How to determine the particular value of these variables depends on the following factors: (1)

`Agent 1`'s own cost and profit structure and evaluation, (2) `Agent 1`'s understanding of the current economic situation and potential demand for its product, (3) `Agent 1`'s model of `Agent 2`, and (4) `Agent 1`'s expectation from `Agent 2`, such as potentially profitable future transactions.

Considering all these factors and the trade-offs among them, `Agent 1` calculates the expected payoff value associated with possible offers, and selects the offer that maximizes his payoff. `Agent 2` receives the offer transmitted by `Agent 1`. To decide whether to accept this offer or to counter-propose, `Agent 2` essentially uses a similar evaluation procedure as `Agent 1`.

The next step would be easy if `Agent 2` decides to accept the offer. In that case, `Agent 2` just needs to send an acceptance message to `Agent 1`, which finalizes the contract. If `Agent 2` is not satisfied with the offer, it can either abort the negotiation or send back a counter-proposal. Again, the process of determining the counter-proposal is similar to that used by `Agent 1` to determine the initial proposal. First, `Agent 2` calculates the payoff function whose domain is all feasible offers. Then, the offer that maximizes `Agent 2`'s payoff is selected. It should be noted that the fact that `Agent 1` has sent a proposal does have an impact on the decision making process that `Agent 2` goes through when deliberating his counter-proposal, since `Agent 2`'s internal knowledge of `Agent 1` and possibly the knowledge about the supply situation have been updated. `Agent 1`'s proposal affects `Agent 2`'s decision in a quite indirect way by causing changes in `Agent 2`'s perception of `Agent 1`.

After `Agent 1` receives the counter-proposal offered by `Agent 2`, `Agent 1` first updates its model of `Agent 2`, then evaluates the offer in the light of newer knowledge. If it is deemed as an acceptable offer, the negotiation process is brought to an end. Otherwise, `Agent 1` sends a counter-proposal based on its payoff structure and newer knowledge about its counterpart, `Agent 2`. Exchanges of proposals and counter-proposals will go on until one of the agents decides to accept an offer or to quit. The negotiation process can also end because of other external events such as missing an agreement deadline, etc.

3.1 Bazaar: a formal description

▶ In Bazaar, a negotiation process can be modeled by a 10-tuple $< N, M, \Delta, A, H, Q, \Omega, P, C, G >$, where,

A-1 A set N (the set of players).

A-2 A set M (the set of issues/dimensions covered in negotiation. For instance, in the supply chain management domain, this set could include product price, product quality, payment method, transportation method, etc.)

A-3 A set of vectors $\Delta \equiv \{(D_j)_{j \in M}\}$ (a set of vectors whose elements describe each and every dimension of an agreement under negotiation).

A set A composed of all the possible actions that can be taken by every member of the players set.

▷ $A \equiv \Delta \cup \{Accept, Quit\}$

A-4 For each player $i \in N$ a set of possible agreements A_i.

▷ For each $i \in N$, $A_i \subset A$.

A-5 A set H of sequences (finite or infinite) that satisfies the following properties:

▷ The elements of each sequence are defined over A.

▷ The empty sequence Φ is a member of H.

▷ If $(a^k)_{k=1,\ldots,K} \in H$ and $L < K$ then $(a^k)_{k=1,\ldots,L} \in H$.

▷ If $(a^k)_{k=1,\ldots,K} \in H$ and $a^K \in \{Accept, Quit\}$ then $a^k \overline{\in} \{Accept, Quit\}$ when $k = 1, \ldots, K-1$.

Each member of H is a *history*; each component of a history is an action taken by a player. A history $(a^k)_{k=1,\ldots,K}$ is terminal if there is no a_{K+1} such that $(a^k)_{k=1,\ldots,K+1} \in H$. The set of terminal histories is denoted by Z.

A-6 A function Q that associates each nonterminal history $(h \in H \setminus Z)$ to a member of N. (Q is the *player function* which determines the orderings of agent responses.)

A-7 A set of Ω of relevant information entities. Ω is introduced to represent the players' knowledge and belief about the following aspects of negotiation:

▷ The parameters of the environment, which can change over time. For example, in supply chain management, global economic or industry-wide indices such as overall product supply and demand and interest rate, belong to Ω.

▷ Beliefs about other players. These beliefs can be approximately decomposed into three categories:

(a) Beliefs about the factual aspects of other agents, such as how their payoff functions are structured, how many resources they have, etc.

(b) Beliefs about the decision making process of other agents. For example, what would be other players' reservation prices.

(c) Beliefs about meta-level issues such as the overall *negotiation style* of other players. Are they tough or compliant? How would they perceive a certain action? What about their risk-taking attitudes? etc.

A-8 For each nonterminal history h and each player $i \in N$, a subjective probability distribution $P_{h,i}$ defined over Ω. This distribution is a concise representation of the knowledge held by each player in each stage of negotiation.

A-9 For each player $i \in N$, each nonterminal history h, and each action $a \in A_i$, there is an implementation cost $C_{i,h,a}$. C can be interpreted as communication costs or costs associated with time caused by delaying terminal action (*Accept* or *Quit*).

A-10 For each terminal history $h = (a^k)_{k=1,\ldots,K}$ and each player $i \in N$, a *preference* relation $\succeq_i$ on h and $P_{h,i}(x)$, $x \in \Omega$. $\succeq_i$ in turn results in an evaluation function $E_X^{(h,i)}[G_i(X,h)]$.

We will present the solution strategy in Bazaar before we discuss the characteristics of the model.

3.2 Solution Strategy in Bazaar

Although the role that the players play (e.g., selling or buying) with respect to initiating the negotiation process can have an impact[2], the decision making process in a negotiation scenario, viz., determining the particular contents of an offer/counteroffer (quit and accept can be viewed as a special offer), is symmetrical for all the players. So the following solution framework is not limited by roles of the players:

- For each player i, a negotiation strategy is a sequence of actions $(a_i^k, k = 1, \ldots, K)$, where
  - k denotes that a_i^k is the kth action ($k \leq K$) taken by i,
  - $a_i^k \in A_i$,
  - $a_i^K \in \{Accept, Quit\}$,
  - $a_i^k \overline{\in} \{Accept, Quit\}$ when $k = 1, \ldots, K-1$
- Before negotiation starts, each player has a certain amount of knowledge about Ω, which may include the knowledge about the environment where the negotiation takes place,

[2]For example, in a 2-player supply chain situation, the supplier often is the first one to initiate a negotiation.

and may also include the prior knowledge about other players (from previous experience or from second-hand knowledge, etc.) This prior knowledge is denoted (see A-8) as $P_{\Phi,i}$.

- Suppose player i has been interacting with another player j for k times. In other words, i has sent exactly k offers or counteroffers to j (presumably received k or $k+1$ offers or counteroffers from j depending on who initiated the negotiation process). Let us assume that neither *Accept* nor *Quit* has appeared in these offers and counteroffers. In Bazaar, the following information is available when i tries to figure out what to do next (the content of its $k+1$th offer):

  1. All the actions taken by all the agents up to the current time point when i makes decision about the $k+1$th offer. Formally, each and every history h that is a sequence of k actions is known to i. Let us denote this set of histories by $H_{i,k}$.
  2. The set of subjective probability distributions over Ω, $P_{H_{i,k-1},i} \equiv \{P_{h,i} | h \in H_{i,k-1}\}$ is known to i.

  The player takes the following steps to decide how to reply to the most recent action taken by other participant(s):

Step 1 Update his subjective evaluation about the environment and other players using Bayesian rules: Given prior distribution $P_{H_{i,k-1},i}$ and newly incoming information $H_{i,k}$, calculate the posterior distribution $P_{H_{i,k},i}$.

Step 2 For $h \in H_{i,k}$, select the best action from A_i according to the following recursive evaluation criteria:

$$V_{i,k,h} = E_X^{(h,i)}[G_i(X,h)] \quad \text{if } h \in Z$$
$$V_{i,k,h} = \max_{a_i \in A_i} \{-C_{i,a_i,h} + \int_X [V_{i,k+1,(h,a_i)} \times P_{h,i}(X)] dX\} \quad \text{otherwise}$$

The first equation represents the termination criterion. The second equation can be summarized as "always choose the action that maximizes the expected payoff given the information available at this stage". The implementation cost C at this stage has been deducted from the future (expected) payoff.

3.3 Characteristics of Bazaar

Most game-theoretic models assume that the player has infinite reasoning and computation capacity. On the one hand, this infinite rationality assumption eliminates some of the theoretical problems (e.g., the precise definition of degree of rationality is unknown) associated with modeling agents with bounded rationality; on the other hand, it is just because of assuming infinite smartness of players that outguessing regress becomes a problem, since every participating agent tries to model others in a recursive fashion (e.g., (Gmytrasiewicz & Durfee 1992)). The fact that the agents do not have infinite reasoning capacity imposes natural termination for otherwise endless outguessing regress. This is precisely the foundation of Bazaar. Bazaar ignores some aspects of the "strategic" part of a game by modeling other players explicitly (see A-7b) in terms of beliefs and uncertainty. This, along with its learning capability, differentiates Bazaar from other negotiation models. Other observations about Bazaar are:

- Bazaar aims at modeling multi-issue negotiation processes. By incorporating multiple dimensions into the action space, Bazaar is able to provide an expressive language to describe the relationships between these issues and possible trade-offs among them.
- Bazaar supports an *open*-world model. Any change in the outside environment, if relevant and perceived by a player, will have an impact on the player's subsequent decision making processes. This feature is highly desirable and is seldom found in other negotiation models.
- In most of existing negotiation models, learning issues have been either simply ignored or oversimplified for theoretical convenience. Multi-agent learning issues can be addressed in Bazaar and conveniently supported by the iterative nature of sequential decision making and the explicit representation of beliefs about other agents.

4 Learning in Negotiation

The importance of learning in negotiation has been recently recognized in the game research community as fundamental for understanding human behavior as well as for developing new solution concepts (Osborne & Rubinstein 1994, Jordan 1992, Kalai & Lehrer 1993). Theoretical results (most of which are partial and preliminary), however, are available only for the simplest game settings. Multi-agent learning has also increasingly drawn research efforts from Distributed AI community (e.g., (Mor, Goldman & Rosenschein 1995, Sen & Sekaran

1995)). In the context of Bazaar, we are using the Bayesian framework to update the knowledge and belief that each agent has about the environment and other agents. To address the computational complexity issues with Bayesian analysis, we use the Bayesian belief network representation and updating mechanism. In addition to providing efficient updating techniques, Bayesian belief networks offer an expressive modeling language and allow easy and flexible encoding of domain-specific knowledge (Pearl 1988).

In this section, we revisit the buyer-supplier example used before to demonstrate how the Bayesian framework can be utilized in a negotiation setting. For illustrative purposes, we consider the negotiation process only from the viewpoint of the buyer and assume that the relevant information set Ω is comprised of only one item: belief about the supplier's reservation price $RP_{supplier}$. An agent's reservation price is the agent's threshold of offer acceptability. Typically a reservation price is private to each agent, and is different for each agent for each negotiation issue. For example, a supplier's reservation price is the price such that the supplier agent will not accept an offer below this price; a buyer's reservation price is the price such that the buyer will not accept an offer above this price. As shown in Figure 1, when the supplier's reservation price $RP_{supplier}$ is lower than the buyer's reservation price RP_{buyer}, any point within the "zone of agreement" is a candidate solution; while, if RP_{buyer} is lower than $RP_{supplier}$, as shown in Figure 2, the zone of agreement does not exist and no deal can be reached via negotiation. If a zone of agreement exists, typically both the buyer and the supplier will make concessions from their initial proposal. The buyer will increase his initial proposal, while the supplier will decrease his. Eventually, a proposal within the zone of agreement will be acceptable to both.

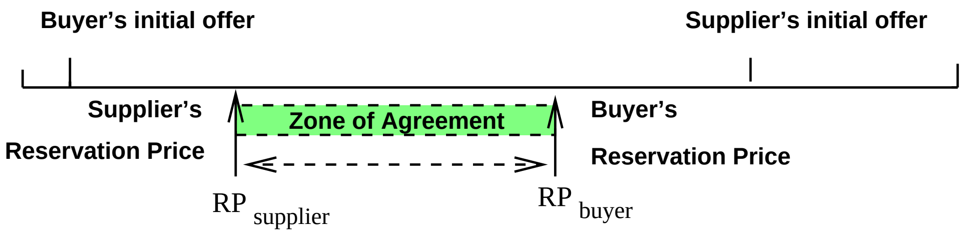

Figure 1: An example of reservation prices and "zone of agreement"

It is obvious that although the buyer knows his own reservation price, the precise value of $RP_{supplier}$ is unknown to him. Therefore, the zone of agreement is not known by either of the agents. Nevertheless, the buyer is able to update his belief (learn) about $RP_{supplier}$ based on his interactions with the supplier and on his domain knowledge. As a result of learning, the buyer is expected to gain more accurate expectation of the supplier's payoff structure and

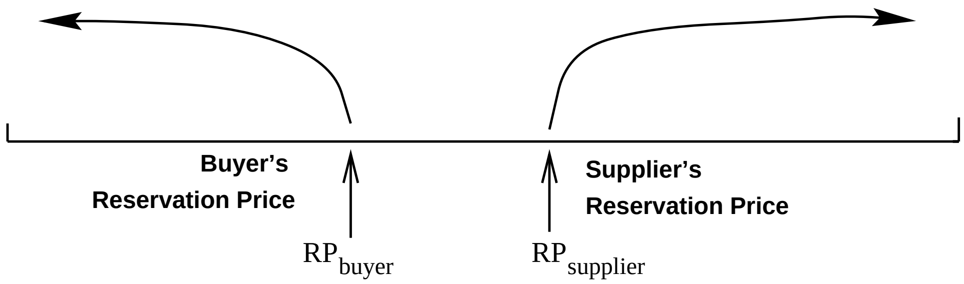

Figure 2: An example in which no "zone of agreement" exists

therefore make more advantageous offers. In this example, we show how the buyer's belief about $RP_{supplier}$ can be updated during negotiation.

The buyer's partial belief about $RP_{supplier}$ can be represented by a set of hypotheses $H_i, i = 1, 2, \ldots, n$. For instance, H_1 can be "$RP_{supplier} = \$100.00$"; H_2 "$RP_{supplier} = \$90.00$". A priori knowledge held by the buyer can be summarized as probabilistic evaluation over the set of hypotheses $\{H_i\}$ (e.g., $P(H_1) = 0.2$, $P(H_2) = 0.35, \ldots$). The Bayesian updating occurs when the buyer receives new signals from the outside environment or from the supplier. Along with domain-specific knowledge, these new signals enable the buyer to acquire new insights about $RP_{supplier}$ in the form of posterior subjective evaluation over H_i. In our case, the offers and counteroffers (Offer$_{supplier}$) from the supplier comprise the incoming signal; while the domain knowledge can be an observation such as: "Usually in our business people will offer a price which is above their reservation price by 17%", which can be represented by a set of conditional statements of similar form, one of which is shown as follows: $P(e_1 \mid H_1) = 0.30$, where e_1 represents "Offer$_{supplier} = \$117.00$", and H_1 "$RP_{supplier} = \$100.00$".

Given the encoded domain knowledge in the form of conditional statements and the signal (e) in the form of offers made by the supplier, the buyer can use the standard Bayesian updating rule to revise his belief about $RP_{supplier}$:

$$P(H_i \mid e) = \frac{P(H_i)P(e \mid H_i)}{\sum_{k=1}^{n} P(e \mid H_k)P(H_k)}$$

We use a numerical example to show how this updating works. For simplicity, we suppose that the buyer knows that the supplier's reservation price is either \$100.00 or \$90.00. In other words, the buyer has only two hypotheses: H_1: "$RP_{supplier} = \$100.00$" and H_2: "$RP_{supplier} = \$90.00$".

At the beginning of the negotiation, the buyer does not have any other additional information. His a priori knowledge can be summarized as: $P(H_1) = 0.5, P(H_2) = 0.5$.

In addition, we suppose that the buyer is aware of "Suppliers will typically offer a price which is above their reservation price by 17%", part of which is encoded as: $P(e_1 \mid H_1) = 0.30$ and $P(e_1 \mid H_2) = 0.05$, where e_1 denotes the event that the supplier asks \$117.00 for the goods under negotiation.

Now suppose that the supplier offers \$117.00 for the product the buyer wants to purchase. Given this signal and the domain knowledge, the buyer can calculate the posterior estimation of $RP_{supplier}$ as follows:

$$\begin{aligned} P(H_1 \mid e_1) &= \frac{P(H_1)P(e_1 \mid H_1)}{P(H_1)P(e_1 \mid H_1) + P(H_2)P(e_1 \mid H_2)} \\ &= 85.7\% \\ P(H_2 \mid e_1) &= \frac{P(H_2)P(e_1 \mid H_2)}{P(H_2)P(e_1 \mid H_1) + P(H_2)P(e_1 \mid H_2)} \\ &= 14.3\% \end{aligned}$$

Suppose that the buyer adopts a simple negotiation strategy: "Propose a price which is equal to the estimated $RP_{supplier}$". Prior to receiving the supplier's offer (\$117.00), the buyer would propose \$95.00 (the mean of the $RP_{supplier}$ subjective distribution). After receiving the offer from the supplier and updating his belief about $RP_{supplier}$, the buyer will propose \$98.57 instead. Since the new offer is calculated based on a more accurate estimation of the supplier's utility structure, it might result in a potentially more beneficial final outcome for the buyer and may also help both sides reach the agreement more efficiently.

Some observations about this example are as follows: (1) Parameters contained in domain knowledge such as the estimated percentage of the supplier's offer over his reservation price (17%) can be updated in a similar fashion. For instance, it is not unrealistic to suppose that this percentage will drop when the negotiation process continues. (2) The belief updating can be triggered by events such as discovery of externally available information in addition to the supplier's offers. For instance, if the buyer finds out during the negotiation the overall supply of the particular goods under negotiation is experiencing a tremendous increase, his estimated supplier's reservation price might drop without even receiving any new offers from the supplier. (3) In this example, we use the traditional Bayesian representation for illustrative purposes. Other efficient updating mechanisms utilizing more expressive representations such as the Bayesian network work essentially in the same way.

4.1 Theoretical Analysis of Utility of Bayesian Learning

In order to examine analytically the impact of learning on negotiation, we make certain simplifying assumptions.

Model and Assumptions A group of n players play an infinitely repeated game. The stage game—the one-shot game being played in a repeated fashion—is described by the following components:

1. n finite sets $\Sigma_1, \Sigma_2, \ldots, \Sigma_n$ of actions with $\Sigma = \prod_{i=1}^{n} \Sigma i$ denoting the set of action combinations.

2. n payoff functions u_i: $\Sigma \mapsto \Re$

We let H_t denote the set of histories of length t, $t = 0, 1, 2, \ldots$). Denote by $\overline{H} = \cup_t H_t$ the set of all finite histories.

A behavior strategy of player i is a function $f_i : \overline{H} \mapsto \Delta(\Sigma_i)$ with $\Delta(\Sigma_i)$ denoting the set of probability distribution on Σ_i. Thus, a strategy specifies how a player randomizes over his choices of actions after every history.

We assume that each player knows his own payoff function and that the players are fully informed about all realized past action combinations at each stage.

The players' objective is to maximize their long term expected discounted payoff, relative to their individual subjective beliefs, including private probabilistic knowledge on the unknown parameters of the game, and beliefs about each others' strategies. Learning takes place at each stage when the players update their individual subjective beliefs using the Bayesian mechanism before entering the next stage of negotiation.

The following definitions are standard game-theoretic concepts (Osborne & Rubinstein 1994),

Nash equilibrium: A Nash-equilibrium of a game is a set of actions with the property that no player can profitably deviate from this equilibrium, given the actions of the other players.

ε-Nash equilibrium: For any $\varepsilon > 0$, an ε-Nash equilibrium of a game is a set of actions with the property that no player has an alternative action that increases his payoff by more than ε, given the actions of the other players.

Lemma 1 *If the players start with a vector of subjectively rational strategies, and if their individual subjective beliefs regarding opponents' strategies are compatible with the truly chosen strategies, then they must converge in finite time to play according to an ε-Nash equilibrium of the repeated game for arbitrarily small ε. (Kalai & Lehrer 1993)*

By subjectively rational strategies we mean that in each stage, the players take the action which maximizes their long term expected discounted payoff, relative to their individual subjective beliefs. Compatibility with the truly chosen strategies means that there should be no event in the play of the infinite game which can occur yet be ruled out by the beliefs of an individual player. Roughly speaking, initially each player assigns a strictly positive probability to the strategy which could be actually chosen by the opponent.

Lemma 2 *After a sufficiently large time T, the real probability distribution over the future play of the game is ε-close to what player i believes the distribution is. (Kalai & Lehrer 1993)*

Based on these two lemmas, we prove the following proposition:

Proposition 1 *A player who uses the Bayesian mechanism to update his beliefs about the unknown parameters of the game and other player's strategies in a subjectively rational fashion performs at least as well as without the Bayesian learning.*

Proof: Suppose a player, P, does not have Bayesian learning capability. P will play without adaptation according to his prior information about other players and unknown parameters of the game. We know from Lemma 2 that all the other players that are able to learn through the Bayesian updating will acquire the almost accurate beliefs over the future play and therefore play according to ε-Nash equilibrium. If P happens to select the *right* strategy at the very beginning, then in the long run learning does not make a difference, since eventually all the agents will play optimally. On the other hand, if P selects a sub-optimal strategy at the outset and cannot adapt its behavior accordingly given the observations of other players' behaviors, it is highly probable that his strategy deviates from ε-Nash equilibrium while others' strategies do not (from Lemma 1). According to the definition of ε-Nash-equilibrium, the proposition immediately follows.

From this proposition, we know that for the simple negotiation setting discussed here, learning is indeed beneficial, which reinforces our intuition that learning helps an agent acquire and update relevant negotiation information during the negotiation process and in turn helps the agent make knowledgeable and advantageous decisions.

5 Experimental Study: Learning in Bargaining

The analytic results given in the previous section ensure the benefit of the Bayesian learning in general. However, the assumptions made by the theory, such as the compatibility of initial subjective beliefs regarding opponent's strategies and the truly chosen strategies, are rarely met in a real negotiation setting. In addition, the theory does not articulate how fast the convergence of the belief update is. The negotiation can be over well before the asymptotic true estimation is achieved.

We conducted simulations in a simple bargaining setting to observe the interactions between the agents that learn and the agents that use fixed strategies. We ran experiments in various situations: learning agents vs. non-learning agents; learning agents vs. learning agents. Results from non-learning vs. non-learning were used as the baseline for comparison.

5.1 Experimental Design

In our initial experiments, we consider a simple bargaining scenario with the following characteristics:

- The set of players N is comprised of one buyer and one supplier
- The set of dimensions M contains only one issue, *price*
- For simplicity, the range of possible prices is from 0 to 100 units
- The set of possible actions (proposed prices by either the buyer or the supplier) A equals to $\{0, 1, 2, \ldots, 100\}$
- The player function Q is defined in such a way that the buyer and the supplier make alternate proposals. Who will be proposing first is decided by coin-tossing
- For simplicity, the relevant information set Ω contains only the supplier's reservation price RP_s and the buyer's reservation price RP_b
- Reservation prices are private information. In other words, each player only knows his own reservation price
- The range of possible prices is public information
- Each player's utility is linear to the final price (a number between 0 and 100) accepted by both players

- Each agent is allowed to propose only strictly monotonically. For example, the supplier's subsequent offers will increase monotonically, while the buyer's offers will decrease monotonically. They are not allowed to propose the same value more than once

Since the bargaining process (proposals/counter-proposals) is symmetrical for the buyer and the supplier, the following discussions about the strategies with or without learning apply to both agents. In our experiments, by the *non-learning agents*, we mean the agent that makes his decision based solely on his own reservation price. For instance, the supplier may start proposing 100 initially. The buyer deems it unacceptable and proposes another value. Since the non-learning supplier does not have a model of the buyer (in terms of the buyer's reservation price), the supplier just compares the buyer's offer with his own reservation price RP_s. If the buyer's offer exceeds RP_s, the supplier will accept the offer and the negotiation process ends. If not, the supplier will propose a value which is below his previous offer by a fixed percentage (in our experiments, the percentage was arbitrarily set to 1.5%) but above RP_s[3]. The non-learning buyer behaves essentially the same way: whenever the supplier's offer is below the buyer's reservation price RP_b, the buyer will accept the offer. Otherwise, the buyer counter proposes by increasing his previous offer by a fixed percentage (again, the proposed value should be below PR_b).

The *learning* agent's negotiation strategy is fundamentally different. Decisions will be made based on both the agent's own and the opponent's reservation price. Note that reservation prices are private information and there is no way that the agent can know the exact value of his opponent's reservation price, even after an agreement has been reached. However, each learning agent can always have some a priori estimation about his opponent's reservation price and update his estimation during the negotiation process using the Bayesian updating mechanism as shown in previous section. In our implementation, the agent represents his subjective beliefs about his opponent's reservation price using a piecewise probability distribution function. This function is implemented as a vector with 101 elements $\vec{\Pi} = [P_0, P_1, \ldots, P_{100}]$[4]. In this vector, P_i represents the agent's current estimation of the probability that his opponent's reservation price is i. The current estimation of his opponent's reservation price itself is calculated as the mean $\sum_{i=0}^{100} i * P_i/101$. In general, the buyer and the supplier will have different initial set of subjective belief vectors $\vec{\Pi}_s^0$ and $\vec{\Pi}_b^0$.

[3]The actual proposed price will be rounded up to an integer value. To satisfy the strict monotonicity assumption about the offers, the minimum difference between the new value and the old one is one unit.

[4]Note that the price range is public knowledge.

The domain knowledge that the learning agents use to update their estimation of their opponent's reservation price is represented by 101 piecewise conditional probability distribution functions $\{DK_i | i = 0, 1, \ldots, 100\}$. The distribution function DK_i essentially expresses what the opponent proposals would look like if his reservation price is i.

In our experiments, these conditional probability functions do not change. We define these functions using a simple heuristic (shown below). In reality, these functions can be learned by repeated exploration of the space of proposals by an agent in repeated negotiation with the same opponent.

The buyer and the supplier have a different set of conditional probability functions. Let us take the buyer's standpoint. One of the conditional distribution function DK_i^b represents the distribution of the possible proposals made by the supplier given that $RP_s = i$. Figure 3 shows the shape of DK_i^b. In essence, this function says that with a very high probability the supplier will propose a value 17% above his true reservation price. Higher or lower than that is less probable. Similar functions are defined for the supplier as well. The difference is that the supplier believes that with a very high probability the buyer will propose a value 17% *below* his true reservation price.

The price range in our experiments was [0..100]. To set up the initial estimation (a priori information), we assumed that the agents do not have information about each other at the beginning and that the agents have an "optimistic" view of each other. The buyer believes that with high probability the supplier's reservation price is 0; while the supplier believes at the beginning of the bargaining that with high probability the buyer's reservation price is 100.

The supplier's negotiation strategy is as follows: given the current estimation of the buyer's reservation price $\widehat{RP_b}$, he will propose a value between RP_s and $\widehat{RP_b}$ equal to $\alpha \times \widehat{RP_b} + (1 - \alpha) \times RP_s$, where α is a parameter set to 0.10 in our experiments. As in the non-learning agent case, the proposed price will be rounded up to an integer value and will be at least one unit less than the previous proposed value. When the counteroffer from the buyer exceeds RP_s, a deal is reached. Otherwise, the supplier updates his estimation of the buyer's reservation price and continues in essentially the same way. The buyer's negotiation strategy mirrors the supplier's.

5.2 Experimental Results

We conducted experiments in three different settings:

1. non-learning buyer vs. non-learning supplier

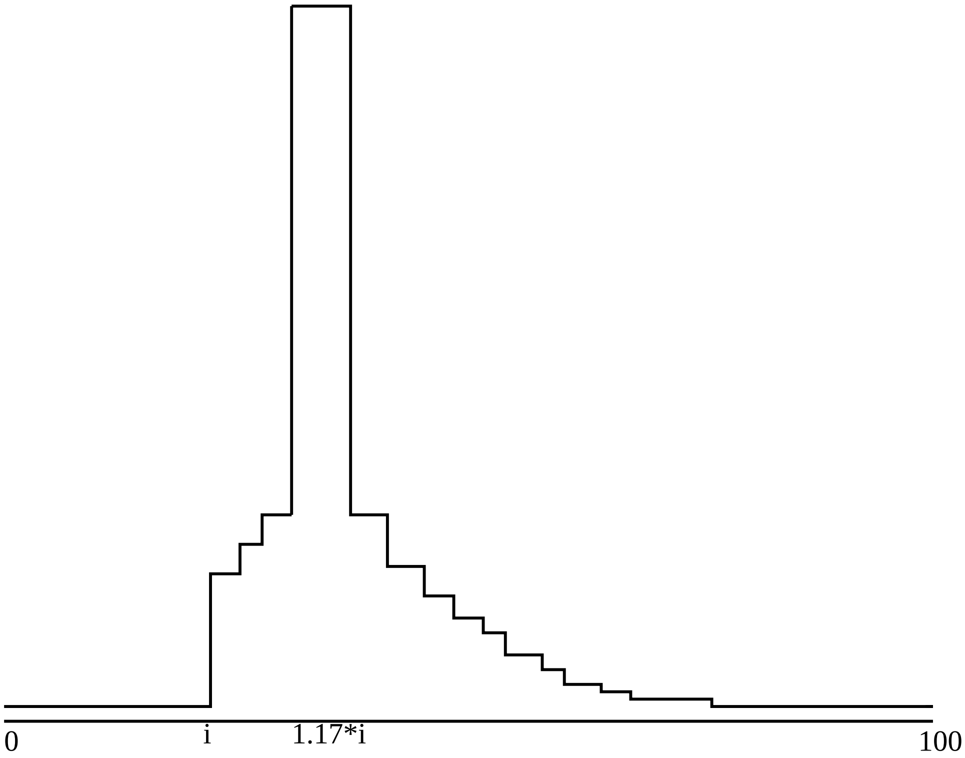

Figure 3: An example of conditional probability function DK_i^b

Configuration	Joint Utility	# of Proposals exchanged
both learn	0.22	24
neither learn	0.18	34
only buyer learns	0.15	28

Table 1: Average Performance of Three Experimental Configurations

2. learning buyer vs. learning supplier
3. learning buyer vs. non-learning supplier

For each configuration, we ran 500 random experiments. Each experiment instance corresponds to a complete bargaining scenario which involves multiple rounds of exchanging proposals and counter-proposals. We generated these 500 random instances of experiment by creating 500 pairs of random numbers. Out of each pair, the lower end, representing the supplier's reservation price, was a realization of a random number that is uniformly distributed in the interval [0..49]. The upper end, representing the buyer's reservation price, was a realization of a random number that is uniformly distributed in the interval [50..100]. In this way, we ensured that the zone of agreement always exists. Note that learning takes place *within* each run of the experiment rather than between the experiment runs.

We measured the quality of a particular bargaining process using the normalized joint utility fashioned after the Nash solution(Luce & Raiffa 1957). Suppose the buyer and the supplier agree on a particular price P_*, the joint utility is then defined as:

$$\frac{(P_* - RP_s) \times (RP_b - P_*)}{(RP_b - RP_s)^2}$$

It can be easily shown that the joint utility reaches the maximum 0.25 when P_* is the arithmetic average of RP_b and RP_s. Note that in our experimental setting this theoretic maximum might not be reached, for RP_b and RP_s are not known to both agents.

The cost of a bargaining process is measured by the number of proposals exchanged before reaching an agreement. We report in Table 1 the average performance of all three configurations. Our observations about these experimental results are as follows.

- We noticed that in terms of overall bargaining quality and number of proposals exchanged to reach a compromise, the "both learn" configuration outperformed the other two. This confirmed our intuition that building learning capability into agents' decision making helps agents form more accurate model of the opponent and results in better performance and less expensive process.

- Judged from the viewpoint of the joint utility, the "only buyer learns" configuration does less well compared with "both learn". In effect, it is even worse than "neither learn". A careful examination of data reveals that although the joint utility suffers, the buyer (the only learning agent) actually did consistently better for himself (in terms of maximizing his own individual utility) than he did in the "both learn" configuration. We suspect the reason is that the buyer has formed better estimation of his non-learning opponent's reservation price and therefore takes advantage of the "dummy" supplier. Since the optimal Nash solution requires an even split in the zone of the agreement, the buyer-dominant solution leads to lower joint utility. The "neither learn" configuration does not show any consistent bias either in favor of the buyer or the supplier.

We examined the data of "neither learn and "both learn" in more detail by further dividing all the 500 experiment instances (1000 instances altogether for both configurations) into different categories according to the size of the zone of agreement. Then, we calculated the differences of the corresponding joint utilities between "neither learn" and "both learn" and plotted the percentage difference in joint utility improvement against the size of the zone of agreement. The result is shown in Figure 4. We observed that there seems to be a positive correlation between these two variables. An intuitive explanation could be that the greater the room for agreement flexibility (greater the zone of agreement), the better the learning agents seize the opportunity.

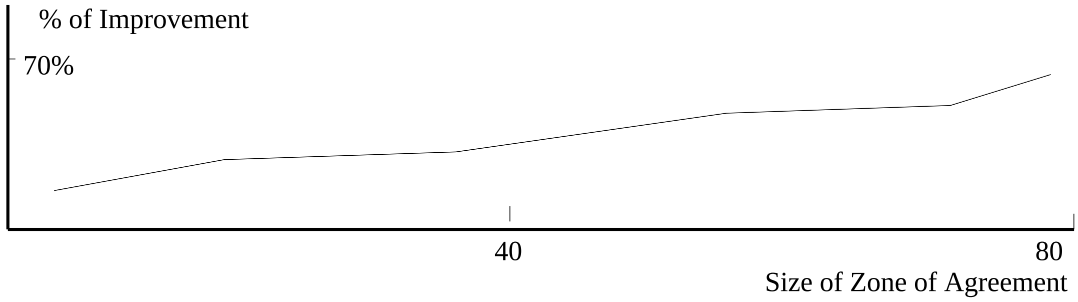

Figure 4: Relations between the Size of the Zone of Agreement & Percent Improvement of Joint Utility

6 Concluding Remarks and Future work

In this paper, we presented Bazaar, a sequential decision making model of negotiation in which multi-agent learning is an integral construct of the model. This model is motivated by providing a computational framework for negotiation which satisfies the following features: (1) the model provides an operational algorithm to guide offers instead of only prescribing the final outcome, and (2) learning can be easily incorporated into the model. Both theoretical results and initial experiments show that learning is beneficial in this sequential negotiation model. Current work focuses on conducting more extensive experiments and theoretical analysis of the impact of learning under various conditions. Future work will investigate the application of the Bazaar framework on nontrivial negotiation scenarios such as supply chain management.

References

Bertsekas, D. P. (1995). *Dynamic Programming and Optimal Control*, Athena Scientific, Belmont, MA.

Cyert, R. M. & DeGroot, M. H. (1987). *Bayesian Analysis and Uncertainty in Economic Theory*, Rowman & Littlefield.

Gmytrasiewicz, P. J. & Durfee, E. H. (1992). Logic of knowledge and belief for recursive modeling: Preliminary report, *Proceedings of the National Conference on Artificial Intelligence*, pp. 628–634.

Harsanyi, J. C. & Selten, R. (1972). A generalized nash solution for two-person bargaining games with incomplete information, *Management Science* **18**: 80–106.

J., R. & Zlotkin, G. (1994). *Rules of Encounter*, MIT Press, Cambridge, Mass.

Jordan, J. S. (1992). The exponential covergence of bayesian learning in normal form games, *Games and Economic Behavior* **4**: 202–217.

Kalai, E. & Lehrer, E. (1993). Rational learning leads to nash equilibrium, *Econometrica* **61**(5): 1019–1045.

Kraus, S. & Subrahmanian, V. S. (1995). Multiagent reasoning with probability, time, and beliefs, *International Journal of Intelligent Systems* **10**(5): 459–499.

Luce, R. D. & Raiffa, H. (1957). *Games and Decisions: Introduction and Critical Survey*, New York, Wiley.

Mor, Y., Goldman, C. V. & Rosenschein, J. S. (1995). Learn your opponent's strategy (in polynomial time), *Proceedings of IJCAI-95 Workshop on Adaptation and Learning in Multiagent Systems*.

Nash, J. F. (1950). The bargaining problem, *Econometrica* **28**: 155–162.

Osborne, M. J. & Rubinstein, A. (1994). *A Course in Game Theory*, The MIT Press.

Pearl, J. (1988). *Proabilistic reasoning in intelligent systems*, second edn, Morgan Kaufmann Publishers, Inc.

Rich, E. & Knight, K. (1991). *Artificial Intelligence*, McGraw-Hill, Inc.

Russell, S. & Wefald, E. (1991). *Do the right thing*, The MIT Press.

Sandholm, T. W. & Lesser, V. R. (1995). Coalition fromation among bounded rational agents, *Proceedings of 14th International Joint Conference on Artificial Intelligence*.

Sen, S. & Sekaran, M. (1995). Multiagent coordination with learning classifier systems, *Proceedings of IJCAI-95 Workshop on Adaptation and Learning in Multiagent Systems*.

Sycara, K. (1990). Negotiation planning: An AI approach, *European Journal of Operational Research* **46**: 216–234.

Young, O. R. (1975). *Bargaining : formal theories of negotiation*, University of Illinois Press.