

# Comparing Equilibria for Game-Theoretic and Evolutionary Bargaining Models

Shaheen Fatima  
Department of Computer  
Science  
University of Liverpool  
Liverpool L69 7ZF, U.K.  
shaheen@csc.liv.ac.uk

Michael Wooldridge  
Department of Computer  
Science  
University of Liverpool  
Liverpool L69 7ZF, U.K.  
mjw@csc.liv.ac.uk

Nicholas R. Jennings  
Department of Electronics and  
Computer Science  
University of Southampton  
Southampton SO17 1BJ, U.K.  
nrj@ecs.soton.ac.uk

## ABSTRACT

Game-theoretic models of bargaining are typically based on the assumption that players have *perfect rationality* and that they always play an equilibrium strategy. In contrast, research in experimental economics shows that in bargaining between human subjects, participants do not always play the equilibrium strategy. Such agents are said to be *boundedly rational*. In playing a game against a boundedly rational opponent, a player's most effective strategy is not the equilibrium strategy, but the one that is the best reply to the opponent's actual strategy. Against this background, this paper studies the bargaining behavior of boundedly rational agents by using genetic algorithms. Since bargaining involves players with different utility functions, we have two subpopulations – one represents the buyer, and the other represents the seller (i.e., the population is asymmetric). We study the competitive co-evolution of strategies in the two subpopulations for an incomplete information setting, and compare the results with those prescribed by game theory. Our analysis leads to two main conclusions. Firstly, our study shows that although each agent in the game-theoretic model has a strategy that is dominant at every period at which it makes a move, the stable state of the evolutionary model does not always match the game-theoretic equilibrium outcome. Secondly, as the players mutually adapt to each other's strategy, the stable outcome depends on the initial population.

## 1. INTRODUCTION

Existing game-theoretic models of bargaining [13, 14, 15] are predicated on the presumption that agents are *perfectly rational*, and that this rationality is *common knowledge*. The participants in these models compute the equilibrium strategy from a theoretical analysis of the game and always play that strategy. In contrast, research in experimental economics [12] suggests that the perfect rationality assumption does not apply in human settings. This research shows that human participants learn how to play games through trial and error, and do not compute the equilibrium from a theoretical analysis of the game. Rather, they experiment with

strategies, observe their payoffs, try other strategies and find their way to a strategy that works well. Such players are said to be *boundedly rational*. This result means game theory cannot always be used as a guide to behavior. An agent's optimal actions may be quite different depending upon whether it is playing against a perfectly rational agent or a boundedly rational person.

This divergence led to the use of evolutionary methods for studying the bargaining behavior of boundedly rational agents [18, 9, 4, 17, 1, 3]. Although for certain games the game-theoretic and evolutionary equilibria coincide [17, 16], in general, it has been shown that the game-theoretic outcome may not always be valid when playing against boundedly rational agents [2]. For instance, [18] and [4] show this in their evolutionary model for the Nash demand game, as do Binmore et al [1] in their evolutionary analysis of Rubinstein's alternating offers game of complete information [13] that has a sub-game perfect equilibrium. Generally speaking, however, this existing work comparing game-theoretic and evolutionary outcomes is based on two main assumptions. Firstly, agents have complete information about the bargaining parameters. Secondly, agents are drawn from a single population, in which all individuals have the same utility function. However, we believe both of these assumptions are unlikely to be true in most practical applications. To rectify this shortcoming, our objective in this paper is to assess to what extent the evolutionary computation of agent strategies matches the game-theoretic results for more realistic scenarios. Thus, we focus not only on an incomplete information setting, but also treat the population as asymmetric.

Specifically, the bargaining behavior of boundedly rational players is studied using genetic algorithms (GAs) in which the population is composed of two separate subpopulations – one representing the buyer and the other representing the seller. We use such asymmetric populations because buyers and sellers have fundamentally different aims and objectives (here represented by different utility functions). Moreover, the buyer and the seller each have time constraints in the form of a deadline and a bargaining cost. In this model, which has been analysed game-theoretically in [7], each agent has a unique strategy that is dominant at every time period at which it makes a move.

In short, the main contribution of this paper is to provide an evolutionary analysis of the above model and compare it with its game-theoretic counterpart. In more detail, our work extends the existing work on comparison of game-theoretic and evolutionary equilibria in the following three ways. Firstly, we analyze games of incomplete information that also have time constraints. Secondly, we study the competitive co-evolution of strategies for asymmetric games. Thirdly, our work also highlights the influence of the initial

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

AAMAS-03 Workshop on Agent-Mediated Electronic Commerce V (AMEC-V) 2003 Melbourne, Australia.

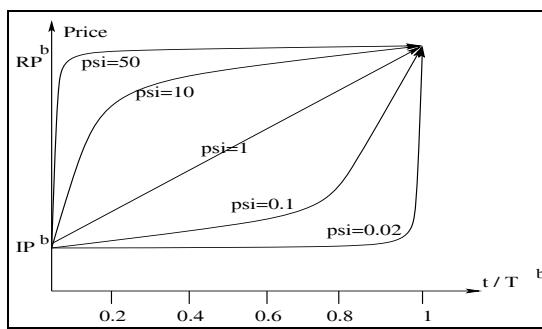


Figure 1: Negotiation decision functions for the buyer.

population on the stable state of the evolutionary model. The results we obtain can then be used to select between the approaches for agent mediated electronic commerce applications, since this decision making involves not only a comparison of the outcomes they generate but also the feasibility of their implementation.

The remainder of the paper is structured as follows. Section 2 gives an overview of our negotiation model. The equilibrium outcomes for this model are presented in Section 3. Section 4 explains the evolutionary system. The results of our experiments are described in Section 5. Section 6 discusses related work and Section 7 gives some conclusions and states our future work.

## 2. THE NEGOTIATION MODEL

We assume that a buyer  $b$  and a seller  $s$  bargain over the price of a good/service. Each agent has an initial price ( $IP$ ) at which it starts negotiation and a reservation price ( $RP$ ) beyond which it does not concede. Let  $[IP^a, RP^a]$  denote the range of values for price that are acceptable to agent  $a$ , where  $a \in \{b, s\}$ ;  $\hat{a}$  denotes agent  $a$ 's opponent. A price that is acceptable to both  $b$  and  $s$ , i.e., the zone of agreement ( $Z$ ), is the interval  $[RP^s, RP^b]$ . The difference between  $RP^b$  and  $RP^s$  is called the *price-surplus*.  $T^a$  denotes agent  $a$ 's deadline. Let  $p_{b \rightarrow s}^t$  denote the price offered by  $b$  to  $s$  at time  $t$ . We use an alternating offers protocol for our study. Negotiation starts when the first offer is made. When an agent, say  $s$ , receives an offer at time  $t$ , i.e.,  $p_{b \rightarrow s}^t$ , it rates the offer using its utility function  $U^s$ . If  $U^s(p_{b \rightarrow s}^t, t)$  is greater than the utility of the counter-offer agent  $s$  is ready to send at time  $t'$ , i.e.,  $p_{s \rightarrow b}^{t'}$  with  $t' = t + 1$ , then agent  $s$  accepts. Otherwise a counter-offer is made. This process of making offers and counter-offers continues until either an agreement is reached, or a deadline is reached.

Since both agents have a deadline, we assume that they use a time dependent tactic (i.e., linear, Boulware, or Conceder [5]) for generating offers. These tactics vary the price depending on the remaining negotiation time. In these functions, the predominant factor used to decide which value to offer next is the time  $t$ . The initial offer is a point in the interval  $[IP^a, RP^a]$ . Agents define a constant  $\gamma^a$ , that, when multiplied by the size of the interval, determines the price to be offered in the first proposal by  $a$ . The offer made by  $a$  at time  $t$  ( $0 < t \leq T^a$ ) is defined in terms of the negotiation decision function (NDF),  $F^a$ , as follows:

$$p_{a \rightarrow \hat{a}}^t = \begin{cases} IP^a + F^a(t)(RP^a - IP^a) & \text{for } b \\ RP^a + (1 - F^a(t))(IP^a - RP^a) & \text{for } s. \end{cases}$$

A wide range of functions can be defined by varying the way in which  $F^a(t)$  is computed. However, functions must ensure that  $0 \leq F^a(t) \leq 1$ ,  $F^a(0) = \gamma^a$  and  $F^a(T^a) = 1$ . That is, the offer

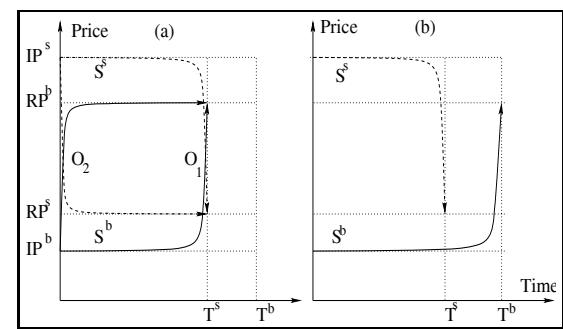


Figure 2: Illustration of agreement and negotiation conflict.

will always be between the value range, at the beginning it will give the initial constant, and when the deadline is reached it will offer the reservation value. The function  $F^a(t)$  is defined as follows:

$$F^a(t) = \gamma^a + (1 - \gamma^a) \left( \frac{\min(t, T^a)}{T^a} \right)^{\frac{1}{\psi}}$$

An infinite number of functions can be defined for different values of  $\psi$ . However, the following two extreme sets show clearly different patterns of behavior [11] (see Figure 1).

1. *Boulware* (B). For this function  $\psi < 1$ , and the initial offer is maintained till time is almost exhausted, when the agent concedes up to its reservation value.
2. *Conceder* (C). For this function  $\psi > 1$ , and the agent goes to its reservation value very quickly<sup>1</sup>. When  $\psi = 1$ , price is increased linearly (L).

The value of a counter-offer depends on the initial price (IP) at which the agent starts negotiation, the final price (FP), beyond which it does not concede,  $\psi$ , and  $T^a$ . A tuple,  $S^a$ , of these four variables, i.e.,  $S^a = \langle IP^a, FP^a, \psi^a, T^a \rangle$ , forms agent  $a$ 's strategy. The *negotiation outcome* ( $O$ ) is an element of  $\{(p, t), \hat{C}\}$ , where  $(p, t)$  denotes the *price* and *time* of agreement and  $\hat{C}$  denotes the conflict outcome.

As an illustration, when  $S^b = \langle IP^b, RP^b, B, T^s \rangle$  and  $S^s = \langle IP^s, RP^s, B, T^s \rangle$ , the outcome ( $O_1$ ) that results is shown in Figure 2. In this figure, and in all subsequent ones, the thick lines denote  $b$ 's strategy and the dashed lines denote  $s$ 's strategy. As shown in the figure, agreement ( $O_1$ ) is reached at a price  $RP^s + (price-surplus/2)$  and at a time close to  $T^s$ . But when the NDF in both strategies is replaced with C, agreement ( $O_2$ ) is reached at the same price but near the beginning of negotiation. Figure 2(b) illustrates a negotiation conflict; where the strategies for  $b$  and  $s$  are  $S^b \times S^s = \langle IP^b, RP^b, B, T^s \rangle \times \langle RP^s, RP^s, B, T^s \rangle$ , and  $T^s < T^b$ . As agents have different deadlines and both agents use the  $B$  function, the strategies do not converge and result in a conflict. In general, an agent can avoid conflict by using a strategy that offers a mutually acceptable price (i.e., within  $Z$ ) by a mutually acceptable time (the earlier deadline).

Agents' utilities are defined with the following two functions that incorporate the effects of bargaining costs:

$$U^a(p, t) = U_p^a(p) + U_t^a(t) = \begin{cases} k_p^b(RP^b - p) + k_t^b t & \text{for } b \\ k_p^s(p - RP^s) + k_t^s t & \text{for } s. \end{cases}$$

<sup>1</sup> As  $\psi$  increases(decreases)  $F$  becomes more Conceder(Boulware). At very high(low) values of  $\psi$ ,  $F$  is an extreme Conceder(Boulware).

For an agent,  $U_t^a$  increases with time if its bargaining cost,  $k_t^a$ , is greater than 0. Consequently, the agent gains utility over time and has the incentive to reach a late agreement. But if  $U_t^a$  decreases with time (i.e.,  $k_t^a < 0$ ), then the agent loses over time and has an incentive to reach an early agreement. Agents are said to have similar time preferences if both gain on time or both lose on time; otherwise they have conflicting time preferences.

An agent's utility from agreement is always higher than its conflict utility. Each agent therefore prefers to reach an agreement rather than disagree and not reach any agreement.

### 3. EQUILIBRIUM OUTCOMES

Each agent has a reservation limit, a deadline, and a bargaining cost. Thus  $b$  and  $s$  each have three parameters denoted  $\langle RP^b, T^b, k_t^b \rangle$  and  $\langle RP^s, T^s, k_t^s \rangle$  respectively. The negotiation outcome depends on all these six parameters. An agent's strategic behavior depends on the information it has about the bargaining parameters. The *information state*,  $I^a$ , of agent  $a$  is the information it has about the negotiation parameters. An agent's own parameters are known to it, but the information it has about the opponent's parameters is not complete. We consider the case where each agent knows its opponent's reservation price, i.e.,  $I^a$  is defined as follows:

$$I^a = \begin{cases} \langle RP^b, T^b, k_p^b, k_t^b, RP^s \rangle & \text{for } b \\ \langle RP^s, T^s, k_p^s, k_t^s, RP^b \rangle & \text{for } s. \end{cases}$$

An agent, say  $b$ 's, optimal strategy depends on the opponent's strategy. Let  $k_t^b > 0$ . As shown in Figure 3, the possible strategies for  $s$  are  $S_1^s$ ,  $S_2^s$ , or  $S_3^s$ . For each of these three  $s$ 's strategies, the strategy that gives  $b$  the maximum utility is  $S_3^b$  (out of  $S_1^b$ ,  $S_2^b$ , and  $S_3^b$ ), since it results in agreement at the lowest price and at the latest time. For the sake of clarity, Figure 3 shows only the extreme Boulware and Conceder functions. However, note that  $S_3^b$  forms  $b$ 's optimal strategy over the entire strategy space that lies between the extreme Boulware and the extreme Conceder. Thus, when  $k_t^b > 0$ ,  $b$ 's optimal strategy is  $\langle RP^s, RP^b, B, T^b \rangle$ . Analogously,  $s$ 's optimal strategy when  $k_t^s > 0$  becomes  $\langle RP^b, RP^s, B, T^s \rangle$ . It has been shown in [7] that the equilibrium strategy profile for the above information state is

$$S_e^b \times S_e^s = \langle RP^s, RP^b, B, T^b \rangle \times \langle RP^b, RP^s, B, T^s \rangle.$$

More specifically,  $S_e^b \times S_e^s$  forms a sequential equilibrium where  $S_e^b$  is  $b$ 's dominant strategy whenever it is  $b$ 's turn to make a move, and  $S_e^s$  is  $s$ 's dominant strategy whenever it is  $s$ 's turn to move. Moreover, this equilibrium is unique. The equilibrium strategies result in an agreement at  $(P, T)$ , where  $P$  denotes  $RP^s$  if  $T^s < T^b$ , and  $RP^b$  if  $T^s > T^b$ .  $T$  denotes the earlier deadline. In other words, the price-surplus goes to the agent with the longer deadline, and an agreement is always reached at the earlier deadline. The equilibrium strategies and the corresponding outcomes for the remaining scenarios (i.e., when  $k_t^b > 0$  and  $k_t^s < 0$ , or  $k_t^b < 0$  and  $k_t^s > 0$ , or  $k_t^b < 0$  and  $k_t^s < 0$ ) are summarised in Table 1 (see [7] for details).

### 4. THE EVOLUTIONARY SYSTEM

The evolutionary model imagines a game as being played not by a single set of players, but by large populations of players. These players are repeatedly and randomly matched to play the game. Each agent is characterized by a strategy that it plays when it is matched. As play proceeds, it observes the payoff of this strategy. It also observes the payoffs and strategies of others (within its population), and has access to information concerning how others have played. In the light of these observations, it adjusts its strategies.

$k_t^b, k_t^s$	$S_{eq}^b \times S_{eq}^s$	Equilibrium Outcome
G,G	$\langle RP^s, RP^b, B, T^b \rangle \times \langle RP^b, RP^s, B, T^s \rangle$	$(P, T)$
G,L	$\langle RP^s, RP^b, B, T^b \rangle \times \langle RP^b, RP^s, C, T^s \rangle$	$(RP^s, T_0)$
L,G	$\langle RP^s, RP^b, C, T^b \rangle \times \langle RP^b, RP^s, B, T^s \rangle$	$(RP^b, T_0)$
L,L	$\langle RP^s, RP^b, C, T^b \rangle \times \langle RP^b, RP^s, C, T^s \rangle$	$(\frac{RP^b + RP^s}{2}, T_0)$

**Table 1: Equilibrium strategies and outcomes for different negotiation scenarios.** *G* indicates  $k_t^a > 0$  and *L* indicates  $k_t^a < 0$ . *P* denotes  $RP^s$  if  $T^s < T^b$ , and  $RP^b$  if  $T^s > T^b$ . *T* denotes the earlier deadline and  $T_0$  denotes the second time period.

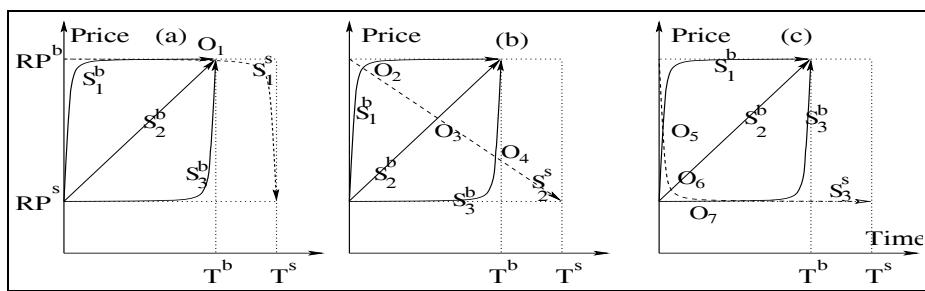
These adjustments involve experimenting with strategies that it has not tried, with the overall aim of switching away from strategies that give low payoffs to strategies that give high payoffs.

Since bargaining involves two agents with different utility functions, we treat the population as being composed of two different subpopulations; one representing the buyer and the other representing the seller. In such asymmetric populations, the evolution of strategies in each subpopulation affects the evolution of strategies in the other subpopulation, (i.e., the strategies co-evolve). Thus we study the competitive co-evolution in which the fitness of an individual in one population is based on direct competition with individuals of the other population.

We represent an agent's fitness with its utility function and apply the three standard operations of *selection*, *crossover*, and *mutation*. An agent's strategy was defined in Section 2 as a tuple of four elements, viz., the initial price, the final price, the negotiation decision function and the deadline. Each individual is represented as a string of fixed length. The bits of the string (the genes) represent the four elements of an agent's strategy. The range of values for these genes are as follows. Since each agent knows its opponent's reservation price (see information state of an agent defined in Section 3), we fix  $IP^a$  to be  $RP^a$ . This is because agreement can never take place outside the zone of agreement. The final price lies in the range  $[RP^b, RP^s]$  (i.e.,  $Z$ ). The NDF can be anywhere between an extreme Boulware and an extreme Conceder for both the agents. The last element is the time at which the final price is offered. Since each agent knows its own deadline, we fix the last element of the strategy at the agent's deadline. In other words, the first and the last elements of the strategy are fixed and do not change. For these two fixed values<sup>2</sup>, the GA needs to find the most effective strategy by varying the *FP* and the NDF. (i.e., the second and third elements of the strategy tuple).

The different stages in an iteration of the GA are as follows. Individuals in the two subpopulations are initialised with some strategies. How the two populations are initialised is explained in Section 5. Once initialised, the parent agents in one of the populations, say the buyer population, start the negotiation process. The fitness of the parent agents in both the populations is determined by competition between the agents in the two populations. Each agent competes against all the agents in the other population. The average utility obtained in these negotiations is then used as the agent's fitness value. In the next stage, offspring agents are created for each

<sup>2</sup>Although the first and last elements of the strategy tuple could be treated as search parameters, we treated them as constants in order to reduce the search space. Note that this encoding includes all possible feasible agreements in the search space, and excludes only that part of the search space where an agreement can never take place (i.e., points outside  $Z$  or beyond the deadline). Consequently the time taken to reach the stable state is reduced.



**Figure 3: Possible buyer and seller strategies when  $k_t^b > 0$  and  $k_t^s > 0$ .**

population using the standard operations of selection, crossover, and mutation. Each of these operations is explained below.

The buyer and seller population size were each set to  $N$ . Selection was carried out using the fitness proportionate selection method [10], where individuals are chosen with a probability proportional to their fitness. To perform crossover within a population, we select two individuals randomly. Two crossover points are then chosen randomly and sorted in ascending order. Then the genes between the successive crossover points are alternately exchanged between the individuals with probability  $P_c$ . Mutation is the process of creating completely new strategies that are not present in the initial population. To perform mutation, a gene (in our case, the second or the third element of the strategy tuple) is selected randomly, and a random value is chosen for it from the domain of the gene. We perform mutation on the second and third elements of the strategy tuple because an optimal value needs to be found for these two elements. The mutation rate was  $P_m$ . We determined the stable outcome for different values of  $N$ ,  $P_c$ , and  $P_m$  in the ranges 20 to 75, 0.1 to 0.9, and 0.005 to 0.05 respectively. Increasing the population size beyond 50 did not change the stable outcome but only increased the time to stabilize. The stable outcome was, on an average, found to be closest to the equilibrium outcome for  $P_c = 0.5$ , and  $P_m = 0.01$ .

To indicate that fitness proportionate selection is a reasonable method in this case, we carried out the above set of experiments using the other common selection method, namely tournament selection [10] with a tournament size of 2. Between the two selection methods, the stable outcome generated by fitness proportionate selection was, on average, found to be closer to the equilibrium outcome. Section 5 therefore describes the evolutionary experiments for the fitness proportionate selection method for  $N = 50$ ,  $P_c = 0.5$ , and  $P_m = 0.01$ .

Note that all genetic operations are carried out within a subpopulation, i.e., there is no transfer of strategies across the two subpopulations. The simulations stop when the population is stable, i.e., 95% of the individuals in each subpopulation have the same fitness, for 10 successive generations. This is because, depending on the initialization of the subpopulations, all the individuals in one or both of the subpopulations can have the same fitness values in the first iteration itself.

## 5. THE EVOLUTIONARY EXPERIMENTS

This section determines the stable outcomes and shows how the initial population affects these outcomes. As mentioned in Section 2, the deadlines are different for  $b$  and  $s$ . Let  $a_l$  denote the agent with the longer deadline and  $a_{sh}$  the one with the shorter deadline. Let  $Pop_l$  and  $Pop_{sh}$  denote the corresponding populations. To determine if the stable outcome depends on the initial

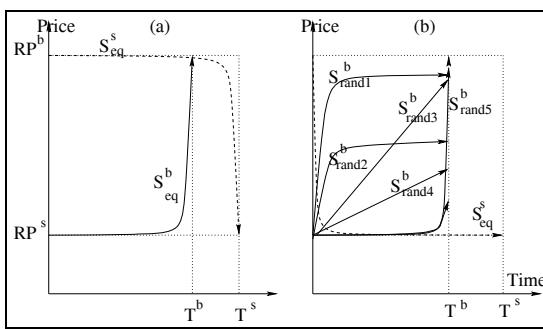
population, we ran the GA for the following different initial populations for each of the four possible negotiation scenarios listed in Table 1.

- $I_1$  Both  $Pop_l$  and  $Pop_{sh}$  are initialised to the game-theoretic equilibrium strategies given in Section 3.
- $I_2$  One of the populations is initialised to the equilibrium strategy and the other to some random non-equilibrium strategies.
- $I_3$  Both  $Pop_l$  and  $Pop_{sh}$  are initialised to some random non-equilibrium strategies.

### 5.1 Both buyer and seller prefer a late agreement ( $k_t^b > 0$ and $k_t^s > 0$ )

The stable state for each of the three initializations is explained below. When all the individuals in each subpopulation are initialised to their respective equilibrium strategies (i.e.,  $I_1$ ), the stable outcome was identical to the game-theoretic equilibrium outcome. As seen in Section 3, in the equilibrium outcome for this scenario, the entire price-surplus goes to the agent with the longer deadline. This evolutionary behavior can be explained by examining how the strategies in the two subpopulations co-evolve. Consider  $Pop_{sh}$  (which represents the buyer) first. Since all the individuals in  $Pop_{sh}$  are initialised to the equilibrium strategy (see Figure 4(a)), they all have the same average fitness values after the first round of negotiations. The other population, i.e.,  $Pop_l$ , is also initialised to its equilibrium strategy, which gives all its individuals the same average fitness values after the first round of negotiations. The new non-equilibrium strategies that get introduced in  $Pop_{sh}$ , due to mutation<sup>3</sup>, either have a lower value for  $FP$  than the  $FP$  in  $S_{eq}^b$ , or an NDF that differs from the NDF in  $S_{eq}^b$ . These strategies mostly conflict with the vast majority of equilibrium strategies of  $Pop_l$  (the non-equilibrium strategies that are subsequently introduced into  $Pop_l$  form a very small fraction of the entire population), resulting in relatively inferior fitness values, and eventually dying out, while the equilibrium strategies, being superior, survive to future generations. Turning now to  $Pop_l$ , the new non-equilibrium strategies that are generated in  $Pop_l$  have a higher value for  $FP$  than the  $FP$  in  $S_{eq}^s$ . In addition, the NDF can be linear, Boulware or Conceder. Those strategies that use a Conceder or linear NDF result in agreement at a lower price, yield a fitness value that is lower than the equilibrium strategy fitness, and as a result do not survive. Those non-equilibrium strategies that use the Boulware NDF result in the same outcome as the equilibrium strategies. In other words, even in  $Pop_l$ , only those individuals that

<sup>3</sup>Since all the individuals play the same strategy, crossover does not yield new strategies. Crossover between two identical strategies results in the same strategy.

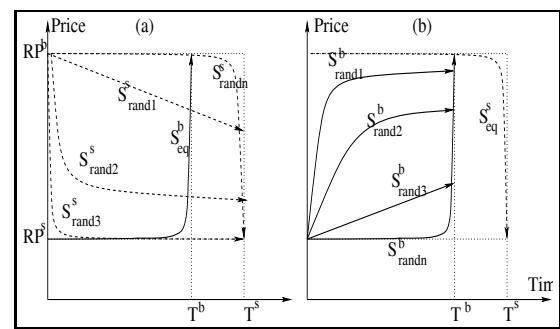


**Figure 4: Buyer and seller population initialization.**

play the equilibrium strategy reach the stable state. The majority of individuals in both the populations continue to play their respective equilibrium strategies. The stable outcome is therefore identical to the game-theoretic equilibrium outcome.

For the second initialization ( $I_2$ ), the stable outcome was found to depend on the initial population corresponding to the agent with the earlier deadline. There are two possibilities for ( $I_2$ ). Either  $Pop_{sh}$  is initialised randomly and  $Pop_l$  with its equilibrium strategy, or  $Pop_{sh}$  is initialised with its equilibrium strategy and  $Pop_l$  randomly. Each of these is explained below. When  $Pop_{sh}$  is initialised with the equilibrium strategy and  $Pop_l$  is initialised randomly, both the populations stabilised at the equilibrium strategies. To understand this, consider  $Pop_{sh}$ . The initial populations for this scenario are depicted in Figure 5(a). Since all the individuals in  $Pop_{sh}$  play the same strategy, they all have the same fitness values. The new strategies that are generated from mutation are non-equilibrium strategies. The individuals that play the equilibrium strategy have a higher fitness than those playing the non-equilibrium strategy. The new strategies introduced from mutation thus get eliminated, while the equilibrium strategy prevails. The other population, i.e.,  $Pop_l$ , is initialised randomly, resulting in a different fitness value to each individual. The closer the strategy is to the equilibrium strategy, the higher its fitness. The close-to-equilibrium strategies thus flourish in  $Pop_l$  at the expense of the non-equilibrium strategies. The behavior of  $Pop_l$  adapts to best suit the predominantly equilibrium strategy of  $Pop_{sh}$ .  $Pop_l$ 's best reply to  $Pop_{sh}$  is the equilibrium strategy. Thus  $Pop_l$  dynamically changes its strategy and stabilizes at the equilibrium strategy. Both populations thus stabilize at the equilibrium strategy.

For ( $I_2$ ), when  $Pop_{sh}$  was initialised randomly, the stable outcome was found to be different from the equilibrium outcome. The agent with the earlier deadline obtains a higher utility than its utility from the equilibrium outcome. But the agent with the longer deadline gets a lower than equilibrium utility. This is explained as follows. Consider  $Pop_l$  first, which is initialised to the equilibrium strategy. The equilibrium strategy of  $Pop_l$  conflicts with most of the strategies of  $Pop_{sh}$ , since they are non-equilibrium (see Figure 5(b)). Moreover, since all individuals play the same strategy, the fitness values are the same for all of them, and correspond to the conflict outcome. The new strategies that are generated from mutation, although being non-equilibrium strategies, result in agreement and are thereby fitter than the equilibrium strategies (recall that an agreement always gives an agent a higher utility than its conflict utility). The number of non-equilibrium strategies thus increases from one generation to the next. But the rate of this change is very slow, since most of the individuals play the equilibrium strategy. The chances that two strategies selected for crossover are



**Figure 5: Buyer and seller population initialization  $I_2$ .**

identical is high. Crossover between two identical strategies yields the same strategy. Thus while mutation can yield a new strategy, the chance of generating new strategies through crossover is low.  $Pop_l$  thus has a very low rate of change. The other population,  $Pop_{sh}$ , is initialised randomly but the fitness levels of the individuals in this population too are equal, and correspond to the conflict outcome. However, as non-equilibrium strategies get generated in  $Pop_l$ , the individuals in  $Pop_{sh}$  have different fitness levels since they play random strategies. Since all individuals play different strategies, the rate of evolution of  $Pop_{sh}$  is faster than  $Pop_l$ . Eventually,  $Pop_l$  evolves towards a strategy that differs slightly from the equilibrium strategy, since only such strategies result in a better agreement with the non-equilibrium strategies of  $Pop_{sh}$ , and yield a higher payoff than the conflict outcome. On the other hand,  $Pop_{sh}$  evolves towards a strategy that is the best reply to this non-equilibrium strategy of  $Pop_l$ . Both populations thus shift away from the initial conflict outcome and eventually stabilize at a non-equilibrium one.

For initialization ( $I_3$ ), where both  $Pop_{sh}$  and  $Pop_l$  are initialised randomly, the stable state was again different from the theoretical equilibrium outcome. It was also different from the stable state for the case where  $Pop_{sh}$  was initialised randomly and  $Pop_l$  with its equilibrium strategy.

The experiments for each of the initialisations  $I_1$ ,  $I_2$ , and  $I_3$  were repeated 50 times. Despite the presence of randomness, we found that the outcomes in these different runs did not vary by more than 3% and the relationship between the outcomes for  $I_1$ ,  $I_2$ , and  $I_3$  always remained the same. These results (averaged over all the runs) are summarised in Table 2. For all the runs,  $RP^b$  was 100,  $RP^s$  was 20,  $k_t^b$  and  $k_t^s$  were both greater than 0,  $T^b$  was 90, and  $T^s$  was 195. As seen in the table, an agreement is always reached at the earlier deadline (i.e.,  $T^b = 90$ ). The price of agreement lies between  $RP^s = 20$  and  $RP^b = 100$  and is close to  $RP^b$ , the reservation price of the agent with the earlier deadline.

Table 2 also shows that the dominant strategy for  $a_{sh}$  (which represents the buyer) is to initialize the population randomly since this results in agreement at a lower price. The dominant strategy for  $a_l$  (which represents the seller) is to initialize all the individuals with its equilibrium strategy. Note that  $b$  and  $s$  have similar time preferences. The time of agreement in all cases was the earlier deadline, which gives the maximum possible utility from time to both the agents. The price of agreement favours  $a_l$ .

To sum up, these experiments show that when each population learns and adapts its behavior to best suit the opponent's behavior, the stable outcome is not always the same as the game-theoretic equilibrium outcome.

	Equilibrium Strategy	Random
Eq. Strategy	(100,90)	(100,90)
Random	(95,90)	(80,90)

**Table 2: Stable outcomes for different initialisations.** Rows indicate  $b$  and columns indicate  $s$ . The first entry in each pair denotes the price, and the second entry the time of agreement.

## 5.2 Both buyer and seller prefer an early agreement ( $k_t^b < 0$ and $k_t^s < 0$ )

The set of experiments described above was repeated for the case where both  $b$  and  $s$  lose utility on time. The stable outcome was the same as the equilibrium outcome, irrespective of whether the two subpopulations were initialised with the equilibrium strategy or randomly. This is explained below. Consider the case where both  $Pop_{sh}$  and  $Pop_l$  are initialised with their respective equilibrium strategies (i.e.,  $I_1$ ). All the individuals have the same fitness levels, and new strategies that get generated from mutation, being relatively inferior, do not reach the stable state. The stable outcome is therefore the same as the equilibrium outcome. The situation where  $Pop_{sh}$  is initialised randomly and  $Pop_l$  is initialised with its equilibrium strategy (i.e.,  $I_2$ ) is depicted in Figure 4(b). As seen in the figure, all the interactions initially result in an agreement. Moreover, the fitness level of all the individuals in  $Pop_l$  is the same, since they all play the equilibrium strategy. The new strategies that are generated by means of mutation, being inferior to the equilibrium strategy, get eliminated. This is because the constant  $k_t^a$  in our utility function is greater than the constant  $k_p^a$  (see Section 2 for the definition of utility function which is used as an individuals fitness). On the other hand, the individuals in  $Pop_{sh}$  have different fitness values, as they all play different random strategies. The closer a strategy to the equilibrium strategy, the higher its fitness.  $Pop_{sh}$  thus evolves to its equilibrium strategy. Both the subpopulations thus stabilize at their respective equilibrium strategies and result in a stable outcome that is identical to the equilibrium outcome.

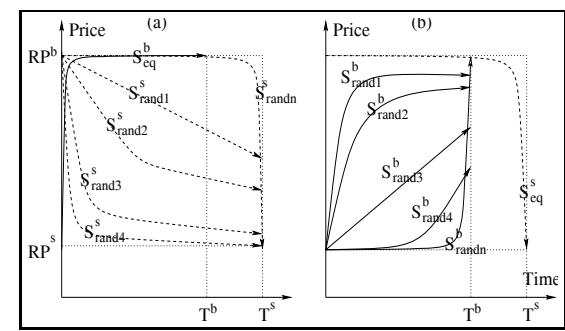
When  $Pop_{sh}$  is initialised with the equilibrium strategy and  $Pop_l$  is initialised randomly, the stable outcome was the same as the equilibrium outcome. As in the previous case, all the interactions initially result in an agreement but,  $Pop_{sh}$  remains stable at the equilibrium strategy, while  $Pop_l$  evolves towards its equilibrium strategy. This co-evolution of strategies eventually results in the same stable outcome as the equilibrium outcome.

Finally, when both  $Pop_{sh}$  and  $Pop_l$  are initialised randomly (i.e.,  $I_3$ ), both populations stabilized at the game-theoretic equilibrium strategy.

To sum up, when  $k_t^b < 0$  and  $k_t^s < 0$ , the stable outcome always matched the equilibrium outcome.

## 5.3 Buyer prefers an early agreement and seller prefers a late one ( $k_t^b < 0$ and $k_t^s > 0$ )

When both  $Pop_{sh}$  and  $Pop_l$  are initialised to their respective equilibrium strategies (i.e.,  $I_1$ ), as in the previous two subsections, the stable outcome was the same as the equilibrium outcome. For  $I_2$ , when  $Pop_{sh}$  is initialised randomly and  $Pop_l$  with its equilibrium strategy, then the stable outcome was the same as the equilibrium outcome. In the initial populations depicted in Figure 6(b), the first round of negotiations mostly result in conflict between the equilibrium strategy of  $Pop_l$  and the random strategies of  $Pop_{sh}$ . Moreover, almost all the individuals in both the populations have the same fitness value, which is equal to the conflict utility. In the next generation, the new strategies that are generated through mutation

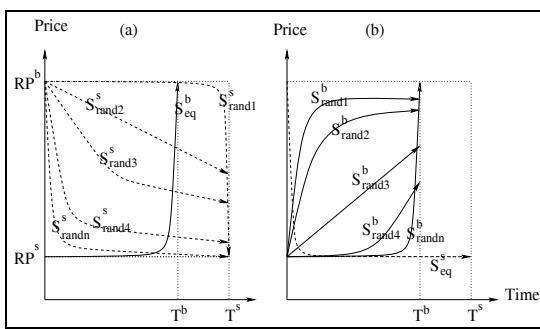


**Figure 6: Buyer and seller population initialization when  $k_t^b < 0$  and  $k_t^s > 0$ .**

in  $Pop_l$ , have a different NDF. They are either less Boulware, linear, or Conceder and result in agreement with some or all of the strategies of  $Pop_{sh}$ . However the majority of the individuals still play the equilibrium strategy, and thereby have the same fitness.  $Pop_l$  thus evolves slowly. Looking at  $Pop_{sh}$  we see that, since all individuals play a different strategy, it evolves relatively faster than  $Pop_l$ . This is because new strategies are generated in  $Pop_{sh}$  through two operations (viz., crossover and mutation) as opposed to the generation of new strategies in  $Pop_l$  through mutation alone. In other words, as  $Pop_{sh}$  evolves faster, it moves towards the strategy that is the most effective reply to the predominantly equilibrium strategy played by  $Pop_l$ . Eventually,  $Pop_{sh}$  stabilizes at its equilibrium strategy and  $Pop_l$  stabilizes at a strategy that is slightly less Boulware than its equilibrium strategy. The stable and equilibrium strategies for  $s$  ( $S_{stab}^s, S_{eq}^s$ ) are depicted in Figure 8(a). As seen in the figure, although the stable strategy of  $Pop_l$  is not the same as the equilibrium strategy, agreement is still reached at the same point as the equilibrium outcome. This is because the difference in price between the two strategies ( $S_{eq}^s$  and  $S_{stab}^s$ ), is high at  $T^b$ , and almost zero near the beginning of negotiation. Since  $Pop_{sh}$  stabilizes at its equilibrium strategy, which uses the Conceder function, the stable outcome results in agreement near the beginning of negotiation and is the same as the equilibrium outcome. Contrast this with the stable outcome corresponding to Figure 5(b), where  $k_t^b > 0$ . Since agreement takes place at  $T^b$ , the stable outcome differs from the equilibrium outcome. But in Figure 8(a),  $k_t^b < 0$ , and the stable agreement takes place at the beginning of negotiation, which is identical to the equilibrium outcome.

For  $I_2$ , when  $Pop_l$  is initialised randomly and  $Pop_{sh}$  with its equilibrium strategy (see Figure 6(a)), the outcome was again the same as the equilibrium outcome. Notice that in this case, initially all the interactions result in agreement since  $Pop_{sh}$  plays the equilibrium strategy (i.e., the Conceder NDF). All the individuals in  $Pop_{sh}$  therefore have the same fitness values. On the other hand, the individuals in  $Pop_l$  have different fitness values.  $Pop_l$  evolves faster than  $Pop_{sh}$ , and stabilizes at a strategy that is the best reply to the equilibrium strategy played by  $Pop_{sh}$ . The new strategies that get introduced in  $Pop_{sh}$ , through mutation, being inferior to its equilibrium strategy, do not survive. Both the populations therefore stabilise at their respective equilibrium strategies, and result in equilibrium outcome.

When both  $Pop_s$  and  $Pop_l$  are initialised randomly (i.e.,  $I_3$ ), the stable outcome was again found to be the same as the equilibrium outcome. Thus, irrespective of the initialization ( $I_1$ ,  $I_2$ , or  $I_3$ ) of the two subpopulations, the stable outcome is always the same as the equilibrium outcome.



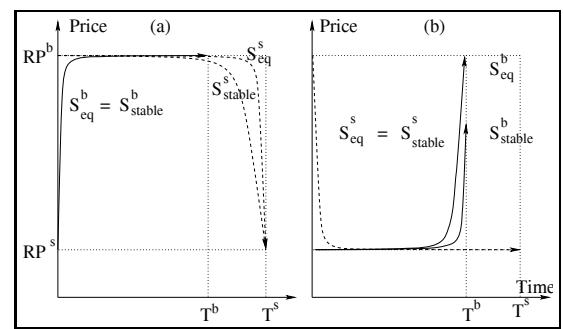
**Figure 7: Buyer and seller population initialization when  $k_t^b > 0$  and  $k_t^s < 0$ .**

#### 5.4 Buyer prefers a late agreement and seller prefers an early one ( $k_t^b > 0$ and $k_t^s < 0$ )

We begin with  $I_1$ , the case where both  $Pop_{sh}$  and  $Pop_l$  are initialised to their respective equilibrium strategies. As in the previous subsections, the stable outcome was the same as the equilibrium outcome. For  $I_2$ , when  $Pop_{sh}$  is initialised randomly and  $Pop_l$  is initialised with its equilibrium strategy, the stable outcome was the same as the equilibrium outcome. Figure 7(b) shows the initial populations for this scenario. As seen in the figure, all the interactions result in agreement. Moreover, since all the individuals in  $Pop_l$  play the equilibrium strategy, they all have the same fitness. The new strategies that are generated through mutation, being inferior to the equilibrium strategy, get eliminated. On the other hand, the individuals in  $Pop_{sh}$  play random strategies and have different fitness values.  $Pop_{sh}$  thus evolves faster than  $Pop_l$ , and stabilizes at a strategy that is the best reply to the equilibrium strategy played by  $Pop_l$ . The best reply to  $Pop_l$  is  $a_{sh}$ 's equilibrium strategy. Both populations thus stabilize at their respective equilibrium strategies. The stable outcome therefore matches the equilibrium outcome.

For  $I_2$ , when  $Pop_{sh}$  is initialised with the equilibrium strategy and  $Pop_l$  is initialised randomly, the stable outcome was again the same as the equilibrium outcome. As shown in Figure 7(a), all the initial interactions between the two populations result in agreement. Furthermore, since  $Pop_{sh}$  is initialised with the equilibrium strategy, all its individuals have the same fitness values, while the individuals in  $Pop_l$  have different fitness values as they are initialised randomly. Within  $Pop_l$ , the closer an individual's strategy is to the equilibrium strategy, the higher its fitness is, since it is the best reply to the equilibrium strategy played by all the individuals of  $Pop_{sh}$ .  $Pop_l$  therefore evolves faster than  $Pop_{sh}$  and stabilizes at its equilibrium strategy. On the other hand, the new non-equilibrium strategies that are generated in  $Pop_{sh}$ , through mutation, have an inferior fitness relative to the equilibrium strategy, and thereby get eliminated. Thus  $Pop_{sh}$  and  $Pop_l$  both stabilize at their respective equilibrium strategies, and result in the equilibrium outcome.

Finally, when both  $Pop_{sh}$  and  $Pop_l$  are initialised randomly (i.e.,  $I_3$ ), the stable outcome was again found to be the same as the equilibrium outcome.  $Pop_l$  stabilized at the non-equilibrium strategy. In the stable strategy of  $Pop_{sh}$  all the elements except the second, (i.e., the final price) were the same as the elements in the equilibrium strategy. Although the final price in the stable strategy was less than  $RP^s$ , it resulted in the equilibrium outcome since the  $IP$ , the NDF and the deadline were the same as in  $S_{eq}^s$  (see Figure 8(b)).



**Figure 8: Stable vs. equilibrium strategies. (a)  $k_t^b < 0$  and  $k_t^s > 0$  (b)  $k_t^b > 0$  and  $k_t^s < 0$ .**

#### 5.5 A summary of key results

The above analysis leads to two main conclusions. Firstly, although game-theoretically each agent has a dominant strategy, the results of our analysis show that when the population is asymmetric and the strategies in the two populations co-evolve, the stable outcome of the evolutionary approach does not always coincide with the game-theoretic equilibrium outcome. Secondly, as the players mutually adapt to each other's strategy, the stable outcome depends on the initial population. More specifically, when  $k_t^b > 0$  and  $k_t^s > 0$ , the outcomes of these two approaches differ. As seen in Table 2, the dominant strategy for  $a_{sh}$  (which in our case represents the buyer) is to initialize the population randomly since this results in agreement at a lower price. The dominant strategy for  $a_l$  (which represents the seller) is to initialize all the individuals with its equilibrium strategy. Also, as shown in Table 2, the outcome generated by the evolutionary approach is more in favour of  $a_{sh}$ , than the equilibrium outcome. The difference however is small. While the equilibrium price of agreement is 100, the price at the stable state (for the right initialization) is 95. For all the remaining scenarios (i.e.,  $k_t^a < 0$  for at least one of the agents) the game-theoretic equilibrium outcome matches the stable outcome of the evolutionary model, irrespective of how the two populations are initialised (i.e.,  $I_1$ ,  $I_2$ , or  $I_3$ ).

From these findings it is clear that implementing software agents using the game-theoretic approach is computationally simpler (since the equilibrium strategies can be determined on the basis of the agents' information states). Once they are determined, the agents just need to be coded with these strategies. In contrast, the stable strategies in the evolutionary model depend on how the populations are initialised. However, an agent may not know exactly how the opponent's population is initialised and, consequently, the GA learning needs to be done online every time there is a negotiation.

## 6. RELATED RESEARCH

A number of game-theoretic models have been studied for the bargaining problem under time constraints [14, 15, 8, 6]. A small, but growing body of literature exists in the field of the application of evolutionary methods to bargaining [9, 18, 4, 1, 3]. Of these, the ones closer to our work are [18, 4, 1, 3]. Young [18] and Ellingsen [4] study the evolutionary model for the Nash demand game, while Binmore et al [1] provide an evolutionary analysis of Rubinstein's alternating offers game of complete information [13] that has a sub-game perfect equilibrium. Cressman and Schlag [3] show the difference between game-theoretic and evolutionary outcomes for extensive form games with distinct payoffs (i.e., games in which no

two paths yield the same payoff for one of the players). However, all these models are based on the assumption that agents are drawn from a single population, in which all the individuals have the same utility function. In a more realistic bargaining scenario, the buyer and the seller have different utility functions. We therefore treat the population as being asymmetric, i.e., the population is composed of two separate subpopulations - one representing the buyer and the other representing the seller since they have different utility functions. Our work thus considers competitive co-evolution, in which fitness is based on direct competition between individuals selected from two independently evolving populations of buyers and sellers. Moreover, we also show that when the two subpopulations co-evolve, then the stable outcome depends on how the two subpopulations are initialised.

Matos et al [9] use GAs to analyse multi-issue negotiation. The population comprises of two subpopulations; one representing the buyer and other representing the seller. They use a fitness function based on the sum of the score across all the competitions. The  $b$  and  $s$  populations evolve simultaneously. In real-life bargaining situations each participant tries to maximize its own utility and not the sum of the participants' utilities. We therefore consider two asymmetric subpopulations in which the strategies co-evolve.

In summary, existing evolutionary models study the bargaining behavior of agents either for a symmetric population (i.e., they assume that both the parties in a game have the same utility function) or study the simultaneous evolution of strategies if the population is asymmetric, by focussing on a specific negotiation scenario. Our work differs from existing models in the following ways. Firstly, we consider an asymmetric population and study the competitive co-evolution of strategies in the two subpopulations. The second difference lies in the stable outcomes generated by the evolutionary models for symmetric and asymmetric populations. For symmetric games of simultaneous offers it has been shown that the evolutionary equilibrium coincides with the Nash equilibrium [17], while for symmetric games of alternating offers the evolutionarily stable outcome are close to the game-theoretic equilibrium under certain conditions [1]. In contrast to this, our study shows that, if the population is asymmetric, the stable outcome of the evolutionary model can differ from the game-theoretic outcome even when each agent has a dominant strategy at every period at which it makes a move. Furthermore, our study also highlights the effect of the initial population on the stable state.

## 7. CONCLUSIONS AND FUTURE WORK

This paper studied the bargaining behavior of boundedly rational agents using GAs and compared the results with the game-theoretic equilibrium outcome, for a particular model of negotiation based on negotiation decision functions. In this negotiation game of incomplete information, each agent has a unique strategy that is dominant at every information state at which it makes a move. In the evolutionary counterpart of this model, there is a competitive co-evolution of strategies between two asymmetric populations. Each player learns the most effective strategy that is the best reply to the opponent's strategy. The key conclusion of our analysis is that although the agents in the game-theoretic model have dominant strategies, the stable state of the corresponding evolutionary model does not always match the equilibrium outcome. Furthermore, as the players mutually adapt to each other's strategy, the stable outcome depends on the initial population.

Our present work used genetic algorithm learning to study the bargaining behavior of boundedly rational agents. In order to get more general results, it would be interesting to extend the analysis to other learning mechanisms.

## Acknowledgements

This research was supported by the EPSRC under grant GR/M07052.

## 8. REFERENCES

- [1] K. Binmore, M. Piccione, and L. Samuelson. Evolutionary stability in alternating-offers bargaining games. *Journal of Economic Theory*, 80:257–291, 1998.
- [2] K. Binmore and L. Samuelson. Evolution and mixed strategies. *Games and Economic Behavior*, 34:200–226, 2001.
- [3] R. Cressman and K. H. Schlag. The dynamic (in)stability of backwards induction. *Journal of Economic Theory*, 83:260–285, 1998.
- [4] T. Ellingsen. The evolution of bargaining behavior. *The Quarterly Journal of Economics*, pages 581–602, May 1997.
- [5] P. Faratin, C. Sierra, and N. R. Jennings. Negotiation decision functions for autonomous agents. *International Journal of Robotics and Autonomous Systems*, 24(3-4):159–182, 1998.
- [6] S. S. Fatima, M. Wooldridge, and N. R. Jennings. An agenda based framework for multi-issue negotiation. *Artificial Intelligence Journal* (to appear).
- [7] S. S. Fatima, M. Wooldridge, and N. R. Jennings. The influence of information on negotiation equilibrium. In J. Padget, O. Shehory, D. Parkes, N. Sadeh, and W. E. Walsh, editors, *Agent Mediated Electronic Commerce IV Designing Mechanisms and Systems*, pages 180–193. Springer Verlag LNAI-2531, 2002.
- [8] S. S. Fatima, M. Wooldridge, and N. R. Jennings. Multi-issue negotiation under time constraints. In *Proceedings of the First International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-02)*, pages 143–150, Bologna, Italy, 2002.
- [9] N. Matos, C. Sierra, and N. R. Jennings. Determining successful negotiation startegies: An evolutionary approach. In *Proceedings of the 3rd International Conference on Multi-Agent Systems (ICMAS-98)*, pages 182–189, 1998.
- [10] M. Mitchell. *An Introduction to Genetic Algorithms*. The MIT Press, Cambridge, Massachusetts, 2001.
- [11] H. Raiffa. *The Art and Science of Negotiation*. Harvard University Press, Cambridge, USA, 1982.
- [12] A. E. Roth. Bargaining experiments. In J. Kagel and A. E. Roth, editors, *Handbook of Experimental Economics*. Princeton NJ, Princeton University Press, 1995.
- [13] A. Rubinstein. Perfect equilibrium in a bargaining model. *Econometrica*, 50(1):97–109, January 1982.
- [14] A. Rubinstein. A bargaining model with incomplete information about time preferences. *Econometrica*, 53:1151–1172, January 1985.
- [15] T. Sandholm and N. Vulkan. Bargaining with deadlines. In *AAAI-99*, pages 44–51, Orlando, FL, 1999.
- [16] J. M. Smith. *Evolution and the theory of games*. Cambridge University Press, Cambridge, 1982.
- [17] J. W. Weibull. *Evolutionary game theory*. MIT Press, 1997.
- [18] H. P. Young. An evolutionary model of bargaining. *Journal of Economic Theory*, 59:145–168, 1993.