Human vs. Computer Behaviour in Multi-Issue Negotiation

Tibor Bosse[1] and Catholijn M. Jonker[2]
Vrije Universiteit Amsterdam, Department of Artificial Intelligence
tbosse@cs.vu.nl, http://www.cs.vu.nl/~tbosse

Radboud Universiteit Nijmegen, Nijmegen Institute for Cognition and Information
C.Jonker@nici.ru.nl, http://www.nici.ru.nl/~catholj

Abstract

This paper presents two experiments that contribute to the comparison of human- versus computer behaviour in the domain of multi-issue negotiation. The experiments are part of an ongoing endeavour of improving the quality of computer negotiators when negotiating against human negotiators. The validity of the experiments was tested in a case study of closed multi-issue negotiation involving the ABMP negotiation software agents. The results indeed reveal a number of strengths and weaknesses of the ABMP agents. For example, the fairness of deals in negotiations performed purely by ABMP agents is better than the fairness of deals in the comparable negotiations in which humans were involved. Furthermore, in mixed negotiations (i.e., involving human- and software agents) the humans outperform the software agent with respect to the individual performance. Based on the results of the experiments, several suggestions are made to improve the ABMP agent's performance.

1. Introduction

Negotiation is an integral part of life, appearing, e.g., at the personal level of individual humans, at the level of companies, and at the level of countries. Over the years there is a steadily increasing stream of papers on the design of software agents for negotiation (see, e.g., [2]). These artificial negotiators are expected to be able to negotiate against other artificial- and against human negotiators. However, negotiation will never be delegated to artificial negotiators (also called agents), if their performance is not at least as good as that of human negotiators.

The need to establish the quality of negotiators implies a need for evaluation tools and experimental setups in which negotiators can be tested against each other. Note that in this formulation, negotiators can be either human or artificial. The SAMIN system for support and analysis of multi-issue negotiation [3] is a software environment that allows negotiators to play against other negotiators and that contains tools to evaluate the negotiation traces against a library of dynamic properties. SAMIN, although still under development, so far is the only such environment.

With respect to experimental research involving negotiation, most literature describes the result of testing different artificial agents against each other (e.g., [5]). Two welcome exceptions are [4] and [10]. In [4], a series of experiments is described where human participants interact with software agents in a Continuous Double Auction. In [10], the performance of a negotiating agent is compared with human performance in the game of Diplomacy. In both papers, the agents are found to outperform the humans. However, the authors of the second paper mention that this might be due to the fact that their participants were not always fully motivated. Moreover, the outcome of both papers is not automatically transferable to other agents and domains. Every new agent will have to prove its worth in a new experiment. The experiment should test the agent with a covering spread of different profiles against human beings also using a covering spread of different profiles. Moreover, note that different domains place different demands on the negotiators. Another contribution to a more rigorous testing of (artificial) negotiation agents is the tournament of [7] in which several agents with different strategies were pitted against each other.

This paper therefore pleads for the development of a benchmark for negotiation that takes into account the different types of negotiation. For an overview of different types of negotiation, see [13]. For every type of negotiation, and for every form of additional constraints (e.g., regarding protocol, time limits and round limits) this benchmark should contain a set of domains with accompanying sets of profiles, a library of properties that should be used to evaluate the negotiations, and a set of experimental setups that should be performed when a new agent is introduced.

Having a standard for testing negotiation agents against humans is especially important. First of all, if the agents are not tested under the same conditions, the results of the tests cannot be compared. Secondly, experiments with humans are time and cost intensive. In our experience, explaining negotiation and the experimental settings to humans approximately takes one hour. On average a negotiation of human against agent takes roughly 30 minutes, a negotiation of human against human takes roughly 60 minutes. Furthermore, the group of humans involved should be sufficiently big to obtain statistically significant results.

If all agents are tested against humans under the same conditions, then the results of those tests can be compared, leading to a comparative analysis of the strengths and weaknesses of the agents involved.

This paper takes a step towards the proposed benchmark by providing the setup of two types of experiments for closed multi-issue negotiation (see, e.g. [13]) and showing their appropriateness by performing them in a case study. The first experiment concerns human-human negotiations, the second concerns human-computer negotiations. The first experiment needs only to be performed once for every domain included in the benchmark. The second needs to be performed for every domain and for every new agent.

The case study concerns a closed multi-issue negotiation on a number of attributes of second hand cars, using the ABMP agent introduced in [8] and the SAMIN system (of [3]) to carry out and analyse the negotiations.

Section 2 describes the formalisation of negotiation process dynamics in terms of negotiation states and traces. In addition, it shows how formal dynamic properties can be specified. The library of dynamic properties that are relevant for closed multi-issue negotiation is presented in Section 3. Section 4 explains briefly how the SAMIN system can use such properties for the analysis of negotiation traces. Next, Section 5 describes the experiments and case study. The results of these experiments for the case study are presented in Section 6. Section 7 completes the paper with a discussion and a description of future research plans.

2. Formalising Negotiation Processes

Negotiation is essentially a dynamic process. To analyse those dynamics, it is, therefore, relevant to formalise and study dynamic properties of such processes. For example, how does a bid at a certain point in time relate to bids at previous time points? The formalisation introduced in this section is based on the notions of negotiation process state and negotiation trace, which are introduced in Section 2.1 and 2.2. Based on these concepts, it is demonstrated in Section 2.3 how formal dynamic properties can be specified.

2.1. Formalising States of a Negotiation Process

The *state* of a (one-to-one) negotiation process at a certain time point can be described as a combined state consisting of two states for each of the negotiating agents: S = < S1, S2 >, where S1 refers to the state of agent A, and S2 to the state of agent B. Each of these states include,

- the agent's own most recent bid
- its evaluation of its own most recent bid
- its evaluation of the other agent's most recent bid
- the history of bids from both sides and evaluations

To describe negotiation states a state ontology Ont is used. Example elements of this ontology are a sort BID for bids, and relations such as util(A, b, v) expressing that A's overall evaluation of bid b is v. Based on this ontology the set of ground atoms At(Ont) can be defined. A state is formalised by a truth assignment: At(Ont) $\rightarrow$ {t, f} to this set of ground atoms. The set of all states described by this ontology is denoted by States(Ont).

2.2. Negotiation Traces

A particular negotiation process shows a sequence of transitions from one state S from States(Ont) to another (next) state S' from States(Ont). A transition S $\rightarrow$ S' from a state S to S' can be classified according to which agents are involved. During such a transition each of the main state components (S1, S2) of the overall state S may change.

Negotiation traces are time-indexed sequences of negotiation states, where each successive pair of states is a negotiation transition. To describe such sequences a fixed *time frame* T is assumed which is linearly ordered. A *trace* γ over a state ontology Ont and time frame T is a mapping γ: T $\rightarrow$ STATES(Ont), i.e., a sequence of states γ_t (t $\in$ T) in STATES(Ont). The set of all traces over state ontology Ont is denoted by TRACES(Ont). Depending on the application, the time frame T may be dense (e.g., the real numbers), or discrete (e.g., the set of integers or natural numbers or a finite initial segment of the natural numbers), or any other form, as long as it has a linear ordering.

2.3. Dynamic Properties

To formally specify dynamic properties that express characteristics of dynamic processes (such as negotiation) from a temporal perspective an expressive language is needed. To this end the *Temporal Trace Language* TTL is used as a tool; cf. [9], which is briefly defined as follows.

The set of *dynamic properties* DYNPROP(Ont) is the set of temporal statements that can be formulated with respect to traces based on the state ontology Ont in the following manner. Given a trace γ over state ontology Ont, a certain state of the agent A during a negotiation process at time point t is indicated by state(γ, t, A). These state indicators can be related to state properties via the formally defined satisfaction relation $\models$, comparable to the Holds-predicate in the Situation Calculus: state(γ, t, A) $\models$ p denotes that state property p holds in trace γ at time t in the state of agent A. Based on these statements, dynamic properties can be formulated in a formal manner in a sorted first-order predicate logic with sorts T for time points, Traces for traces and F for state formulae, using quantifiers and the usual first-order logical connectives such as $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\forall$, $\exists$.

As an example, consider the idea of making concession steps, which is a necessary action to take in order to move towards agreement (see, e.g., [5], [12]). A concession step can be expressed by the following dynamic property: "in trace γ, agent A makes a concession step between time points t1 and t2, if it makes bid b1 at t1 and it makes its next bid b2 at t2, and for agent A bid b2 has a lower utility than bid b1, while for the other agent B it has a higher utility". In the TTL language, this property can be formulated as follows:

concession_step(γ:TRACE, t1:time, t2:time, A:AGENT, B:AGENT) ≡
∃b1, b2:BID
  state(γ, t1, A) |= to_be_communicated_to_by(b1, B, A) &
  state(γ, t2, A) |= to_be_communicated_to_by(b2, B, A) &
  agent_consecutively_bids_to(γ, A, t1, b1, B, t2, b2) &
  ∀vA1, vA2, vB1, vB2: real :
    util(A,b1,vA1) & util(A,b2,vA2) & util(B,b1,vB1) & util(B,b2,vB2)
    ⇒ vA1 > vA2 & vB1 < vB2

where util(A, b1, vA1) expresses that the utility of negotiator A with respect to bid b1 is the value vA1. See [13] for a definition of utility in negotiations. In the rest of the paper the formalisations will be kept to a minimum.

3. Properties of Negotiation Processes

To analyse the differences between human and computer negotiations, two categories of properties of the negotiations are investigated: those that concern the negotiators' *performance* in the negotiation, and those that concern the *steps* in their bidding behaviour. All properties are part of the current library for analysing negotiations. The library contains more properties, omitted here for reasons of space, see [3]. To improve readability, all properties are provided in an informal (natural language) notation, instead of the formal (TTL) notation introduced above.

3.1. Performance Properties

To measure the performance of the different parties in the negotiation, a number of different properties from the literature (e.g., [13], [14]) are included in the library:

- *Negotiator Final Utility*: a number between 0 and 1, indicating the negotiator's utility for the final bid in the negotiation (i.e., the bid that both parties agreed upon). The higher the utility, the higher the satisfaction of the negotiator. A high (but < 1) number does not mean that the negotiator could not have performed better. Furthermore, the utility of the one does not give any information about the utility of the other.
- *Pareto Distance*: a number between 0 and $\sqrt{2}$, indicating the shortest distance from the final bid in the negotiation to the *Pareto Efficient Frontier*. The Pareto Efficient Frontier (PEF) is the set of bids for which there exists no better bid for both parties, see e.g., [13]. Let x_i (respectively y_i) be the utility of negotiator X (respectively Y) for bid b_i. The distance d(b1, b2) between bids b1 and b2 is defined $\sqrt{((x1-y1)^2+(x2-y2)^2)}$. Thus, a short Pareto Distance means that there was little room for the negotiators to improve the outcome for both parties. However, this does not give much information about the fairness of the outcome.
- *Nash Distance*: a number between 0 and $\sqrt{2}$, indicating the distance from the final bid in the negotiation to the *Nash Point* (i.e., the point for which the product of both parties' utilities is maximal, see e.g., [13]). Since the Nash Point lies on the Pareto Efficient Frontier, a short Nash distance implies a short Pareto distance. The Nash Point is considered a fair outcome.
- *EPP Distance*: a number between 0 and $\sqrt{2}$, indicating the distance from the final bid in the negotiation to the *Equal Proportion of Potential Point* (also called the

Kalai-Smorodinski Point, i.e., the point for which the difference between both parties' utilities is minimal, see e.g., [13]). Since the EPP Point also lies on the Pareto Efficient Frontier, a short EPP distance implies a short Pareto distance as well. The EPP Point is also considered a fair outcome.

- *Number of rounds*: a natural number, indicating the number of rounds the negotiation process took. One round consists of a bid made by the seller, followed by a bid made by the buyer. The smaller this number, the quicker an agreement was reached.

3.2. Step Properties

Besides observing the quality of the outcome, the trajectory of bids offered by each of the negotiators is of interest. Each trajectory is composed of the steps made by the negotiator. Every step satisfies exactly one of the following properties that are inspired by [3]:

- *Fortunate steps*: the next bid is better for yourself and better for the other agent
- *Concession steps*: the next bid is worse for yourself and better for the other agent
- *Selfish steps*: the next bid is better for yourself and worse for the other agent
- *Unfortunate steps*: the next bid is worse for yourself and worse for the other agent

As argued in [6], agreement in multi-issue negotiation can often be reached quicker if both parties make concessions. In the experiments described in the next section, it will be investigated to what extent the different types of steps are used, both by human and computer negotiators.

4. SAMIN: The System for Analysis of Multi-Issue Negotiation

To carry out and analyse a number of experiments in negotiation (see next section), the SAMIN system by [3] was used. This Section briefly explains the working of the system.

At the top level, SAMIN consists of three components: an Acquisition Component, an Analysis Component and a Presentation Component, see Figure 1 Here, the solid arrows indicate data flow. The dotted arrows indicate that each component can be controlled separately by the user. The Acquisition Component is used to acquire the input necessary for analysis. The Analysis Component is used to perform the actual analysis (i.e., checking which properties hold for the negotiation process under analysis). Finally, the Presentation Component is used to present the results of the analysis in a user-friendly format. Furthermore, SAMIN maintains a library of properties, templates of properties, bid ontologies, and profile ontologies (not shown in Figure 1. The working of the three components will be described briefly in the next subsections. For more details, see [3].

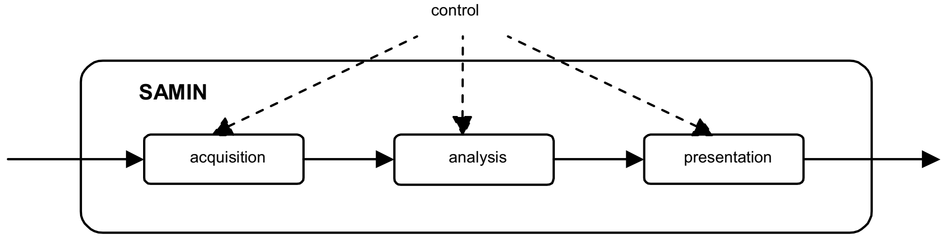

Figure 1. Global overview of the SAMIN architecture

4.1. The Acquisition Component

The acquisition component is used to obtain the required input for the analysis. It consists of an *ontology editor*, a *dynamic property editor* and a *trace determinator*.

The ontology editor is used for the construction of bid ontologies and profile ontologies necessary to automatically interpret the bids exchanged by the negotiators, and to automatically interpret the profiles of the negotiators. The ontology editor is typically used to construct a bid ontology and a profile ontology, thus allowing the user to identify the issues to be negotiated, the values that each of these issues can take, and the structure of bids, in the bid ontology. A *profile* is a description of the preferences of the negotiator within the particular negotiation domain. Thus, in specifying the profile ontology the user identifies the possible evaluations that can be given to values, and the utility functions of bids.

The dynamic property editor supports the gradual formalisation of dynamic properties in TTL format. The editor offers a user interface that allows the analyst to construct dynamic properties, represented in a tree-like format.

The trace determinator can be used interactively with the analyst to determine what traces to use in the analysis. The user can interactively locate the files containing the traces to be checked. The traces themselves can be of three categories: (human) empirical traces, simulated traces, and mixed traces. An empirical trace is the result of an existing human negotiation process. A simulated trace is the result of an automated negotiation process. A mixed trace is the result of a human negotiating with a software agent. To support the acquisition of traces of all three types, a dedicated interface has been created for SAMIN.

4.2. The Analysis Component

The analysis component currently consists of a *logical analyser* that is capable of checking properties against traces. To this end, the tool takes a dynamic property in TTL format and one or more traces as input, and checks whether the dynamic property holds for the traces.

Traces are represented by sets of Prolog facts of the form *holds(state(m1, t(2)), a, true)* where m1 is the trace name, t(2) time point 2, and a is a state property as introduced in Section 2.1. The above example indicates that state formula a is true in trace m1 at time point 2. The Analysis Component basically uses Prolog rules for the predicate sat that reduce the satisfaction of the temporal formula finally to the satisfaction of atomic state formulae at certain time points, which can be read from the trace representation. Examples of such reduction rules are:

```
sat(and(F,G)) :- sat(F), sat(G).
sat(not(and(F,G))) :- sat(or(not(F), not(G))).
sat(or(F,G)) :- sat(F).
sat(or(F,G)) :- sat(G).
sat(not(or(F,G))) :- sat(and(not(F), not(G))).
```

In addition, if a dynamic property does not hold in a trace, then the software reports the places in the trace where the property failed.

4.3. The Presentation Component

The presentation component currently includes a tool that visualises the negotiation space in terms of the utilities of both negotiators. This *visualisation tool* plots the bid trajectory in a 2-dimensional plane, see Figure 2. The utilities are real values that indicate how a particular bid is evaluated by a negotiator.

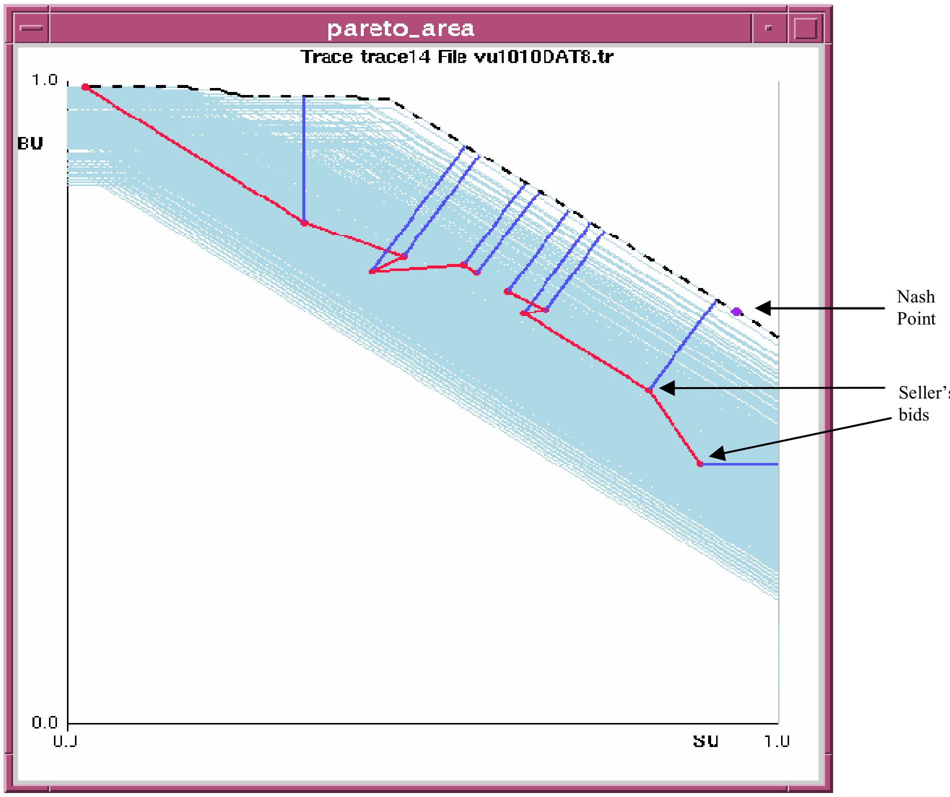

Figure 2. SAMIN Visualisation Tool

In Figure 2, the seller's utility of a bid is on the horizontal axis, and the buyer's utility is on the vertical axis. The light area corresponds to the space of possible bids. In this area, each curve is a continuous line, corresponding to a different combination of discrete issues. The specific position on the line is determined by the continuous issue 'price'. Since in this particular domain 4 discrete issues with 5 possible values occur (see next section), there are already 625 (= 5^4) different curves. In this figure, the sequences of actual bids made by both buyer (left) and seller (right) are indicated by the dark points that are connected by the two angular lines. The upper-left point indicates the buyer's first bid, and the lower-right point indicates the seller's first bid. The dotted line indicates the Pareto Efficient Frontier according to the profiles of the negotiating agents, and the short dark lines show the distance from each bid to this frontier. The small dot that is plotted on the Pareto Efficient Frontier corresponds to the Nash Point. From this picture, it is clear that both negotiators make more and more concessions (their bids converge towards each other). Eventually, they reach a point that does not lie on the Pareto Efficient Frontier, but is rather close to it anyhow.

5. The Experiments

Pre-experiments with 10 participants showed that a Human-Human (HH) negotiation for multi-issue negotiations for non-trivial domains takes approximately one hour to complete. The subjects showed signs of fatigue, and when asked to perform another negotiation, they showed a lack of motivation. The negotiations were performed much quicker, but the results were obviously sub-optimal. On the basis of these observations, the experiments proposed in this paper are limited to one negotiation per human.

Sections 5.1 and 5.2 describe the setup of the experiments. The case study, a negotiation about second hand cars involving the ABMP agents and the SAMIN system, is presented in Section 5.3. The results of the analysis of the acquired traces are presented in Section 6. The strategy used by the ABMP agent can be summarised by the following steps (see [8] for details):

1. For each negotiation round, determine evaluations of the attributes of the previous bids.
2. Aggregate these evaluations into overall utilities of these previous bids.
3. Determine which concession step will be made for the next bid, expressed in terms of the overall utility; this provides a target utility.
4. To obtain the next bid, given the target utility, determine target attribute evaluation values, according to some distribution over attributes (chosen in such a manner that they aggregate exactly to the target utility)
5. For each of these target attribute evaluation values, choose an attribute value that has an evaluation value as close as possible to the target evaluation value for the attribute.

Since this strategy is based on the idea of monotonic concession, we hypothesized that the computer negotiator mainly uses concession steps. On the contrary, human negotiators probably are more diverse in their behaviour.

5.1. Experiment 1: HH

Participants. Gather a selection of humans representative of the adult population. The selection should contain enough humans to possibly gain statistically significant results. The size of the group depends on the number of variables in the domain.

For the case study eighteen subjects participated in this experiment. The make up of the group was not representative of the adult population in general, but was representative of the population of AI students in The Netherlands. The group consisted of 12 males and 6 females. All participants were students in AI, their age varying between 19 and 27 years.

Preparation. Before starting the experiment, the participants are to be provided enough background information to be able to perform the negotiation and use the software environment used to register the negotiations. The participants should be motivated to do their best during the negotiation.

In the case study the participants were motivated by the challenge to obtain a high utility, and to perform better than the computer in the corresponding Computer-Computer negotiation process (CC) they were also allowed to perform. The participants formed 9 groups of two persons, and each group was assigned to a computer.

Method. Each group has to participate in two negotiation processes: a HH process and a CC process. In the HH process, one person is assigned the role of the buyer, and the other one is assigned the role of the seller. The human buyer negotiates with the human seller (both using their own profile). The subjects are not allowed to look at the screen while the other party makes a bid. In the CC process, a computer buyer negotiates with a computer seller (both using the profile of the corresponding human negotiator). By keeping the negotiation profile stable over the two processes, it is guaranteed that the utility spaces remains the same, and that the resulting traces are thus comparable.

5.2. Experiment 2: HC

Participants. Gather a selection of humans representative of the adult population. The selection should contain enough humans to possibly gain statistically significant results.

In the case study 76 subjects (43 males and 33 females) participated in this experiment. The experiment took place during an introductory course for family members of AI students. Most of the participants (about 75%) were parents of the students, their age varying between 45 and 55 years. The other 25% were brothers and sisters of the students, their age varying between 17 and 24 years. Almost all of the participants did not have any background in AI. Education and occupation were a fair representation of the general population in The Netherlands.

Preparation. Before starting the experiment, the participants are to be provided enough background information to be able to perform the negotiation and use the software environment used to register the negotiations.

In the case study the game theoretic notions were treated a bit less thoroughly than in the case study for Experiment 1. The participants formed 38 teams of two persons, and each team was assigned to a computer. Each team was told that they could negotiate as a team against the computer. This deviation was necessary for that occasion, due to a lack of available computers.

Method. Each group has to participate in two negotiation processes: a Human-Computer (HC) process and a CC process. In the HC process, all teams play the same role (e.g., the buyer role), and use their own personal profile. In the CC process, a computer buyer uses the profile of the human team. By keeping the negotiation profile stable over the two processes, it is guaranteed that the utility spaces remain the same, and that the resulting traces are thus comparable.

5.3. Case Study

The case study concerned a multi-issue closed negotiation on second hand cars. To support and analyse the experiments, the SAMIN system [3] was used.

The object of negotiation in the case study is a particular second hand car, for which the relevant issues are cd_player, extra_speakers, airco, drawing_hook and price. Consequently, a bid consists of an indication of which CD player is meant, which extra speakers, airco and drawing hook, and what the price of the bid is. The goal of the negotiators is to find agreement upon the values of the four accessories and the price. Here, the price issue has a continuous value, whilst the other four issues have a value from discrete sets.

Before the negotiation starts, both parties specify their *negotiation profile*, see [8]. SAMIN offers negotiators a graphical interface to specify their personal profile. In addition to this negotiation profile, the seller is also provided with a *financial profile*, describing for each issue how much it costs, both to buy it and to build it into the car. Since we focus on closed negotiation, none of the profiles will be available for the other negotiator. However, SAMIN has access to both profiles.

During the negotiation, all subjects could input their bids within a special interface that also shows the history of bids. To help human negotiators generating their bids, the system offers a special tool that allows the player to calculate the utility of a bid before passing it to the other party. The resulting negotiation traces were logged by the system, so that they could be used for the purpose of analysis.

6. Results

Using the SAMIN system, the properties for multi-issue negotiation introduced in Section 3 have been automatically checked against the traces that resulted from the experiments. This section shows the results of the analysis. Section 6.1 focuses on Experiment 1, and Section 6.2 focuses on Experiment 2. Each section distinguishes between the properties concerning the parties' performance, and those concerning their bidding behaviour. Section 6.3 discusses the most important results of both experiments.

6.1. Experiment 1

6.1.1. Performance Properties: The results with respect to the performance of the negotiators in Experiment 1 are shown in Table 1. The first row contains the mean outcomes over all 9 HH traces. The second row contains the mean outcomes over all 9 CC traces. To test whether the differences between these two means were significant, paired t-tests have been performed, of which the results are shown in the last two rows. For example, the first column states that in the HH traces, the mean utility of the buyer was 0.87, that in the CC traces, the mean utility of the buyer was 0.88, but that this difference was not significant ($t=0.38$, $p<0.717$).

Table 1. Performance in Experiment 1

	Buyer Utility	Seller Utility	Pareto Distance	Nash Distance	EPP Distance	Number of rounds
HH traces	0.87	0.80	0.05	0.22	0.16	7.00
CC traces	0.88	0.89	0.03	0.12	0.06	8.00
t-value	0.376	2.807	-0.786	-3.988	-3.463	1.540
p-value	0.717	0.023	0.455	0.004	0.009	0.146

As can be seen in the table, the results in the second, fourth and fifth column are significant. Thus, the following conclusions can safely be drawn from the experiments:

- the seller's mean utility was significantly higher in the CC traces than in the HH traces ($t=2.81$, $p<0.023$)
- the mean Nash Distance was significantly shorter in the CC traces than in the HH traces ($t=-3.99$, $p<0.004$)
- the mean EPP Distance was significantly shorter in the CC traces than in the HH traces ($t=-3.46$, $p<0.009$)
- with respect to the other properties, there was no significant difference between the HH traces and the CC traces

6.1.2. Step Properties: The results with respect to the step properties in Experiment 1 are shown in Table 2. The numbers between brackets are the percentages with respect to the total amounts of steps. The first row shows the steps made by the (human) buyers in the HH traces. The second row shows the steps made by the (human) sellers in the HH traces. Similarly, the third and fourth row show the steps made by the computer buyers, respectively sellers in the CC traces. For example, the first cell indicates that in the HH traces, all (human) buyers together made 3 fortunate steps, which is 6.52% of the total amount of steps they made.

This table clearly shows that both the human and computer negotiators primarily made concession steps. Besides that, the computers made some unfortunate steps, more than the humans did. The humans were more diverse in their behaviour, since they also made some selfish and some fortunate steps.

Table 2. Steps made in Experiment 1

	Fortunate (S+ O+)	Concession (S- O+)	Selfish (S+ O-)	Unfortunate (S- O-)
HH, buyer	3 (6.52%)	36 (78.26%)	2 (4.35%)	5 (10.87%)
HH, seller	5 (11.36%)	32 (72.73%)	4 (9.09%)	3 (6.82%)
CC, buyer	0 (0%)	58 (89.23%)	0 (0 %)	7 (10.77 %)
CC, seller	0 (0 %)	48 (82.76%)	0 (0 %)	10 (17.24 %)

6.2. Experiment 2

6.2.1. Performance Properties: The results with respect to the performance of the negotiators in Experiment 2 are shown in Table 3. The first row indicates the mean outcomes over all 38 HC traces, the second row indicates the mean outcomes over all 38 CC traces, and the last two rows show the results of the paired t-tests.

Table 3. Performance in Experiment 2

	Buyer Utility	Seller Utility	Pareto Distance	Nash Distance	EPP Distance	Number of rounds
HC traces	0.89	0.72	0.05	0.30	0.23	8.84
CC traces	0.87	0.83	0.06	0.17	0.10	8.91
t-value	-1.729	3.684	0.309	-5.161	-6.228	0.066
p-value	0.092	0.001	0.759	0.000	0.000	0.948

The results in the second, fourth and fifth column are significant. Thus, the following conclusions can safely be drawn from the experiments:

- the seller's mean utility was significantly higher in the CC traces than in the HC traces ($t=3.68$, $p<0.001$)
- the mean Nash Distance was significantly shorter in the CC traces than in the HC traces ($t=-5.16$, $p<0.000$)
- the mean EPP Distance was significantly shorter in the CC traces than in the HC traces ($t=-6.23$, $p<0.000$)
- with respect to the other properties, there was no significant difference between the HC traces and the CC traces

6.2.2. Step Properties: The results with respect to the step properties in Experiment 2 are shown in Table 4. This table shows the same trends as Table 2. Again, both the human and computer negotiators primarily made concession steps. Besides that, the computers made some fortunate and some unfortunate steps, and no selfish steps. The humans, on the other hand, were more diverse in their behaviour. They made significantly more steps in the non-concession categories than the computer.

Table 4. Steps made in Experiment 2

	Fortunate (S+ O+)	Concession (S- O+)	Selfish (S+ O-)	Unfortunate (S- O-)
HC, buyer	23 (7.62%)	232 (76.82%)	17 (5.63%)	30 (9.93%)
HC, seller	2 (0.68%)	251 (85.37%)	0 (0 %)	41 (13.95%)
CC, buyer	0 (0 %)	287 (94.41 %)	0 (0 %)	17 (5.59 %)
CC, seller	0 (0 %)	267 (90.51 %)	0 (0 %)	28 (9.49 %)

6.3. Discussion

Since the profiles used in Experiment 1 differ from those in Experiment 2, it makes no sense to compare the exact data from both experiments with each other. To illustrate this, suppose that in Experiment 1 the profiles of buyer and seller were generally much more similar than in Experiment 2. In that case, in Experiment 1 it would be much easier for both parties to obtain high utility values. Nevertheless, it is possible to compare the general trends one can observe in both experiments.

One trend observed in both experiments, is that the Nash distance and the EPP distance (both measures for fairness of the negotiation) were very short in the CC traces. Table 1 shows that these distances were significantly shorter in the CC traces than in the HH traces, and Table 3 shows that they were shorter in the CC traces than in the HC traces. Furthermore, these distances seem to be shorter in the HH traces than in the HC traces. Thus, the CC negotiations turned out to have the “fairest” outcome, followed by the HH traces. The outcomes of the HC traces were the least balanced. This can be seen in the first two cells in Table 3, where the mean (human) buyer utility (0.89) was much higher than the mean (computer) seller utility (0.72). This is an important finding, because when the same negotiation spaces are explored by two computer negotiators, the buyer utility hardly drops (0.87), whilst the seller utility increases significantly (0.83). Apparently the computer seller is not robust to being exploited by a human buyer. This observation is supported by the data in Table 2 and 4. In both situations, the computers made more unfortunate steps than the humans. In addition, the computer sellers made more unfortunate steps than the computer buyers. A detailed analysis of the traces revealed that the computer seller mostly makes these unfortunate steps when it raises the price of the bid in order to compensate for a better component. Better results might be obtained by making this price compensation a bit lower.

Another explanation of the fact that CC negotiations seem to reach fairer outcomes than HH and HC traces might come from the similarity of the ABMP buyer and seller agents. Both agents follow the same concession strategy, and the parameters that can be used to tune and vary the behaviour of the ABMP agents, were the same in both agents. This explanation can be tested by running the CC negotiations again with different parameter settings for both agents.

A last important finding concerns the diverse bidding behaviour by humans. As shown in Table 2, human negotiators sometimes make steps that improve the utility for both parties. Of course, doing this has the risk of making selfish steps. In its current state, the ABMP agent hardly makes these kinds of steps. Nevertheless, in some cases the unpredictable human behaviour actually resulted in better results. Therefore, it could be beneficial to incorporate these strategies in the software agent as well. Furthermore, it might even help to introduce some (seemingly) unfortunate steps. In many CC traces, the parties found agreement upon the "discrete" issues very quickly, leaving only the price to negotiate upon. However, introducing more unfortunate steps (and thereby changing the values of the discrete issues) might help to escape from "local optima", in the same manner as simulated annealing techniques in evolutionary computing do. A similar idea to improve performance is to consider so-called *weak* concessions: the negotiator changes some values while keeping his utility stable.

7. Conclusion and Future Work

This paper pleads for a benchmark for negotiations. The benchmark is argued to be essential for the improvement of artificial negotiators. The benchmark should provide a general framework and software environment for the comparison of the negotiators with human and artificial negotiators. The work of [3], [4], [7], and [10] can be seen as earlier contributions to such a benchmark.

The two experimental setups presented in this paper were shown to contribute to that benchmark as well. The experiments focus on the comparison of human with computer behaviour in the domain of one-to-one multi-issue negotiation. To validate the use of the experimental setups, a case study was performed. In the first experiment, human-human negotiation was compared with computer-computer negotiation. In the second experiment, human-computer negotiation was compared with computer-computer negotiation. Both experiments yielded a number of interesting results, which can be used to improve the quality of software negotiators in the future.

In fact, these results show that at a number of points the agent of the case study already outperforms the humans. These points mainly involve the fairness of the negotiating outcome. However, with respect to the individual performance (i.e., without caring about the utility of the other party) there is still some room for improvement. Based on the results of the experiments, some suggestions have been made to improve the agent of the case study. Examples are decreasing the price compensations by the seller, introducing more switches between issues, and introducing weak concessions.

With respect to related work, the existing literature describes several other systems that aim at the formal analysis of negotiation processes. Most of these systems focus on CC negotiation, not on HC, CH or HH negotiation. Nevertheless, a number of papers also mention dynamic properties of negotiation processes, e.g., [11], [14], [15]. Like in the current paper, a number of these properties are oriented towards the outcome of the negotiation, in terms of Pareto distance or Nash distance. In addition to this, our paper also identifies a number of properties that are geared towards the *dynamics* of the negotiation process instead of the outcome. In return, in [11], [14], [15] some additional properties are mentioned that are oriented towards rationality and use of resources.

For future research, it is planned to develop and perform more experiments. In particular, a series of experiments will be developed and performed where in all possible combinations (CC, HH, HC, CH) the same profiles are used. Here, especially the CH case (computer buyer vs. human seller) is interesting. Another plan is to extend the SAMIN framework to ease experiments in other domains and open the system for any artificial agent. From the current results, it is still debatable to what extent the conclusions drawn may be generalised. Although the particular strategies of the ABMP agents were designed to be representative for most agent-based strategies in multi-issue negotiation, more experiments with other software agents (or with ABMP agents using different parameter settings) would be welcome.

Finally, from a psychological perspective, it is planned to perform a "Turing test" for negotiation. Following the ideas of [1], a tournament may be set up involving a number of human and automated negotiators. In such a setting, the human participants will have to find out whether they are dealing with an automated or a human negotiation partner.

Acknowledgements

The authors are grateful to Lourens van der Meij for his contribution to the development of the software environment, and to all participants in the case study.

References

[1] Arifovic, J. (2005). *The Implementation of the Turing Tournament: A Report*. Simon Fraser University, Burnaby, Canada. Technical Report.

[2] Beam, C. and Segev, A. (1997). Automated negotiations: a survey of the state of the art. *Wirtschaftsinformatik*, 39(3), 1997, pp. 263–268.

[3] Bosse, T., Jonker, C.M., and Treur, J. (2004). Experiments in Human Multi-Issue Negotiation: Analysis and Support. In: Jennings, N.R., Sierra, C., Sonenberg, L., and Tambe, M. (eds.), *Proceedings of the Third International Joint Conference on Autonomous Agents and Multi-Agent Systems, AAMAS'04*. IEEE Computer Society Press, 2004, pp. 672-679.

[4] Das, R., Hanson, J.E., Kephart, J.O., and Tesauro, G. (2001). Agent-Human Interactions in the Continuous Double Auction. In: *Proceedings of the 17th International Joint Conference on Artificial Intelligence, IJCAI'01*. Morgan Kaufman, 2001.

[5] Faratin, P., Sierra, C., and Jennings, N.R. (1998). Negotiation decision functions for autonomous agents. In: *International Journal of Robotics and Autonomous Systems*, vol. 24(3-4), 1998, pp. 159 - 182.

[6] Gutman, R. and Maes, P. (1998). Agent-mediated Integrative Negotiation for Retail Electronic Commerce. In: P. Noriega and C. Sierra (eds.), *Agent Mediated Electronic Commerce*, Lecture Notes in AI, vol. 1571, 1998, pp. 70-90.

[7] Henderson, P., Walters, B., Crouch, S. and Ni, Q. (2003). A comparison of some negotiation algorithms, in Agent Technologies, Infrastructure, Tools and Applications for E-Services, Springer-Verlag, Lecture Notes in AI, vol. 2592, 2003, pp. 137-150.

[8] Jonker, C.M. and Treur, J. (2001). An Agent Architecture for Multi-Attribute Negotiation. In: B. Nebel (ed.), *Proceedings of the 17th International Joint Conference on AI, IJCAI'01*, 2001, pp. 1195 - 1201.

[9] Jonker, C.M. and Treur, J. (2002). Compositional Verification of Multi-Agent Systems: a Formal Analysis of Pro-activeness and Reactiveness. *International Journal of Cooperative Information Systems*, vol. 11, 2002, pp. 51-92.

[10] Kraus, S. and Lehmann, D.J. (1995). Designing and building a negotiating automated agent. *Computational Intelligence*, 11(1), 1995, pp. 132-171.

[11] Lomuscio, A.R., Wooldridge, M., and Jennings. N.R. (2000). A classification scheme for negotiation in electronic commerce, In: *International Journal of Group Decision and Negotiation*, vol. 12(1), January 2003.

[12] Pruitt, D.G. (1981). *Negotiation Behavior*, Academic Press.

[13] Raiffa, H. (1996). *Lectures on Negotiation Analysis*, PON Books, Program on Negotiation at Harvard Law School, 513 Pound Hall, Harvard Law School, Cambridge, Mass. 02138, 1996.

[14] Rosenschein, J.S. and Zlotkin, G. (1994). *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers*. The MIT Press, Cambridge, MA, 1994.

[15] Sandholm, T. (1999). Distributed rational decision making. In: Weiss, G., (ed.), *Multi-agent Systems: A Modern Introduction to Distributed Artificial Intelligence*, MIT Press, 1999, pp. 201 – 258.

0-7695-2480-X/05 $20.00 © 2005 IEEE