From axiomatic to strategic models of bargaining with logical beliefs and goals

Quoc Bao Vo and Minyi Li
Faculty of Information & Communication Technologies
Swinburne University of Technology, Australia
email: {bvo | myli}@swin.edu.au

ABSTRACT

In this paper, we introduce axiomatic and strategic models for bargaining and investigate the link between the two. Bargaining situations are described in propositional logic while the agents' preferences over the outcomes are expressed as ordinal preferences. Our main contribution is an axiomatic theory of bargaining. We propose a bargaining solution based on the well-known egalitarian social welfare for bargaining problems in which the agents' logical beliefs specify their bottom lines. We prove that the proposed solution is uniquely identified by a set of axioms. We further present a model of bargaining based on argumentation frameworks with the view to develop a strategic model of bargaining using the concept of minimal concession strategy in argument-based negotiation frameworks.

Categories and Subject Descriptors

I.2.11 [**Distributed Artificial Intelligence**]: *Multiagent Systems, Intelligent Agents*; J.4 [**Computer Applications**]: *Social and Behavioral Sciences*

General Terms

Design, Economics

Keywords

Bargaining and negotiation, Collective decision making, Judgment aggregation and belief merging

1. INTRODUCTION

The formal theory of bargaining originated with John Nash's seminal papers [10, 11]. Nash's 1950 paper establishes a framework for bargaining analysis. In this paper, Nash initiated an *axiomatic approach* to bargaining, in which we abstract from the bargaining process itself and specify a list of properties (*axioms*) that a bargaining solution should satisfy. Nash then proposed four axioms and proved that they uniquely characterise what is now known as the Nash bargaining solution. In Nash's 1953 paper, he then turned to the question of how this solution might be obtained in bargaining situations between self-interested agents; i.e., investigate the bargaining problem using a strategic approach. In this paper, Nash implicitly established a new research agenda, attempting to utilise

the strategic (non-cooperative) approach to provide the foundations for cooperative bargaining solution concepts.[1] His approach was to design a non-cooperative game, now known as the Nash demand game, in which the only equilibrium outcome is exactly the one suggested by Nash solution.

We'll now turn our attention to multiagent systems in which agents hold beliefs about the environment they are operating in and have goals they want to achieve or maintain. In many multiagent frameworks, the agents beliefs and goals are represented as logical sentences. Moreover, the agents are required to interact, coordinate and in most cases, negotiate to reach agreements about who do what and who get what. Given that Nash's theories and most literature on bargaining have been based exclusively on utility theory (and possibly probability theory when uncertainty is present), it has been a challenge to apply these game-theoretic models to develop solutions for or analyse the agent negotiation problems when agents hold logical beliefs and goals. Attempts have been made to convert agents' goals to a form of utility through a cost function (see e.g., [13]). However, questions remain on how agents' logical beliefs can be integrated into such a framework. Other researchers have attempted to revive Nash's approaches, particularly the axiomatic approach, by studying the properties that a bargaining solution (of a bargaining problem with logical goals) should satisfy (see e.g., [9, 8, 16, 15] and the reference therein). While [9] and [8] study a number of logical properties for negotiation, it's not clear what bargaining solution they would suggest for self-interested bargainers. In [16], an interesting bargaining solution is proposed together with a study on a number of game-theoretic properties the proposed solution satisfies. Zhang [15] introduces a framework that is perhaps closest to what Nash intended with his axiomatic approach. In this paper, Zhang proposes a solution, which he calls *simultaneous concession solution*, and shows that it is exactly characterised by a set of axioms. Zhang's bargaining solution is, however, quite problematic because: (i) it's syntax sensitive and, as a consequence, prone to manipulation; and (ii) it actually removes the goals that both agents agree on, the so-called "drowning effect".

The present paper is an effort to reopen the Nash program, particularly for agent-based bargaining problems in which agents' beliefs and goals are expressed as logical sentences. Towards that end, we introduce axiomatic and strategic models of bargaining. We propose a set of axioms that a bargaining solution should satisfy and introduce a bargaining solution that is exactly characterised by the proposed axioms. Our proposed bargaining solution is quite intuitive and based on the well-known egalitarian social welfare. Subsequently, we also present a strategic model of bargaining based on the minimal concession strategy and shown that its equilibrium

---

[1]This research agenda has been commonly referred to as the Nash program (see [2]).

outcomes turns out to be the solution outcomes described by our axiomatic theory.

The paper is organised as follows: we present some technical definitions in Section 2. The axiomatic model of bargaining is introduced in Section 3. In particular, we consider two cases: when the agents' bottom lines are fixed and based entirely on the agents' initial beliefs; and when the agents' bottom lines can be revised as the negotiation progresses and the opponent can introduce convincing arguments to challenge the agent's initial bottom line. A main result is also introduced in Section 3. Section 4 presents a strategic bargaining model which is based on the minimal concession strategy with some substantial modification to allow the agents to hold their position without having to make a concession through the use of sufficiently convincing arguments. We follow the anonymous reviewers' recommendations by: (i) omitting a number of preliminary results in Section 4, replacing them instead by a discussion about our model and solution; and (ii) providing the proofs for all of the theoretical results present in this paper. We agree with the reviewers that the revised paper is more self-contained and thus, significantly improved.

2. BACKGROUND

2.1 Logical Preliminaries

We consider a propositional language $\mathcal{L}$ defined from a finite (and non-empty) alphabet $\mathcal{P}$ together with the standard logical connectives, including the Boolean constants $\top$ and $\bot$. Furthermore, we also assume that $\mathcal{P}_O \subseteq \mathcal{P}$ is the non-empty alphabet for the negotiation outcomes. That is, the propositional variables from $\mathcal{P}_O$ constitute the issues to be settled by the negotiating agents. An interpretation ω is a total function from $\mathcal{P}$ to $\{\top, \bot\}$. An interpretation ω is a model of a set of sentences $\Phi \subseteq \mathcal{L}$ if and only if every sentence in Φ is satisfied by ω. $\llbracket \Phi \rrbracket$ denotes the set of models of the set of $\mathcal{L}$-sentences Φ. Given a sentence $\phi \in \mathcal{L}$, we'll also write $\llbracket \phi \rrbracket$ instead of $\llbracket \{\phi\} \rrbracket$. An outcome (or alternative) o is a total function from $\mathcal{P}_O$ to $\{\top, \bot\}$. We denote by $\mathcal{PO}$ the set of all possible outcomes. We will identify each outcome o with the canonical term on $\mathcal{P}_O$ which has o as its unique model. For instance, if $\mathcal{P}_O = \{p, q\}$ and $o(p) = \top, o(q) = \bot$, then o is identified with the term $p \wedge \neg q$ (or $p\bar{q}$, or $\{p, \neg q\}$).

2.2 Nash bargaining theory

Nash [10] established a framework to study bargaining. In his framework, a set of bargainers $N = \{1, 2\}$ tries to come to an agreement over a set of possible alternatives A. If they fail to reach an agreement, a disagreement event D occurs. Each agent $i \in N$ has a von Neumann - Morgenstern utility function $u_i : A \cup \{D\} \rightarrow \mathbb{R}$ from which the set of all utility pairs that result from some agreement,

$$\mathcal{S} = \{(u_1(a), u_2(a)) \in \mathbb{R}^2 \ : \ a \in A\},$$

as well as the pair $d = (d_1, d_2)$, where $d_i = u_i(D)$ can be constructed.

Nash then defined the pair $(\mathcal{S}, d)$ to be a bargaining problem. Nash subsequently defined the bargaining solution to be a function $f : \mathcal{B} \rightarrow \mathbb{R}^2$ that specifies, for each bargaining problem $(\mathcal{S}, d)$, a unique outcome $f(\mathcal{S}, d) \in \mathcal{S}$.

In the same paper, Nash also introduced an axiomatic theory of bargaining. Rather than specifying an explicit model of the bargaining procedure, the axiomatic approach aims to impose properties that one wants a bargaining solution to satisfy, and then look for solutions with these properties. Nash proposed the following four axioms:

A1. *Invariance to equivalent utility representations.* Let the bargaining problem $(\mathcal{S}', d')$ be obtained from $(\mathcal{S}, d)$ by the transformations $s_i' = \alpha_i s_i + \beta_i$ and $d_i' = \alpha_i d_i + \beta_i$, where $\alpha_i > 0$, then $f_i(\mathcal{S}', d') = \alpha_i f_i(\mathcal{S}, d) + \beta_i$, for $i = 1, 2$.

A2. *Symmetry.* If the bargaining problem $(\mathcal{S}, d)$ is symmetric (i.e., $d_1 = d_2$ and $(s_1, s_2) \in \mathcal{S} \Leftrightarrow (s_2, s_1) \in \mathcal{S}$), then $f_1(\mathcal{S}, d) = f_2(\mathcal{S}, d)$.

A3. *Independence of irrelevant alternatives.* If $(\mathcal{S}, d)$ and $(\mathcal{S}', d)$ are bargaining problems such that $\mathcal{S} \subseteq \mathcal{S}'$ and $f(\mathcal{S}', d) \in \mathcal{S}$ then $f(\mathcal{S}', d) = f(\mathcal{S}, d)$.

A4. *Pareto efficiency.* If $(\mathcal{S}, d)$ is a bargaining problem with $s, s' \in \mathcal{S}$ and $s_i' > s_i$ for $i = 1, 2$, then $f(\mathcal{S}, d) \neq s$.

Another important contribution in Nash's seminal paper is that, he also tied the axiomatic theory of bargaining and his proposed bargaining solution up nicely by proving that the proposed axioms uniquely characterise the Nash bargaining solution.

3. A LOGICAL MODEL OF BARGAINING

Consider a finite set $N = \{1, 2\}$ of two agents who try to come to an agreement over the alternatives in $\mathcal{O} \subseteq \mathcal{PO}$. Each agent i has a preference relation $\succeq_i$, which is assumed to be a total pre-order (i.e., total, reflexive and transitive), defined over the set of alternatives $\mathcal{O}$. Each agent i also maintains a set of beliefs $B_i \subseteq \mathcal{L}$. Essentially, B_i represents agent i's beliefs about her available options outside of the negotiation or simply her reservation value.

REMARK: It's important to note that the agents' beliefs don't encode their hard constraints. If an agent has a hard constraint that can never be violated and that rules out an outcome $o \in \mathcal{PO}$ then $o \notin \mathcal{O}$. Rather, the agent's beliefs B_i encode her bottom line in the bargaining, in the sense that, due to the requirement of individual rationality, she will never agree on an outcome that is worse than her bottom line.

We now define the bargaining problem to be a tuple $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$, where $\mathcal{O} \subseteq \mathcal{PO}$ is the set of outcomes and B_i and $\succeq_i$ are agent i's beliefs and preference relation over $\mathcal{O}$, respectively. Following Nash [11] and other researchers in the literature of (cooperative) game-theoretic bargaining (see e.g., [12]), we are also interested in a *bargaining solution*, by which we mean a function f that specifies a unique outcome set $f(BP) \in 2^{\mathcal{O}}$ for every bargaining problem BP. The reason why we target an outcome set as the solution of the bargaining problem instead of an outcome in $\mathcal{O}$ will become clear later.

EXAMPLE 1. *A vendor agent of a house (agent 1) is negotiating with a prospective buyer (agent 2) over the sale of the house. The two issues they need to come to an agreement are: the price of the house (i.e. whether the buyer should pay the vendor's asking price), a proposition denoted by P, and the settlement (i.e., whether it should be an early settlement), denoted by E. Then $\mathcal{O} = \{PE, P\bar{E}, \bar{P}E, \bar{P}\bar{E}\}$ is the set of possible outcomes. In this example, an outcome, say $\bar{P}\bar{E}$, indicates that the buyer would pay a price lower than the vendor's asking price such as the median house price of the area and the settlement will be according to the standard settlement of three months after the date of purchase. Assume further that F denotes the proposition that the vendor agent has an existing offer agreeing to pay him the asking price, and A denotes the proposition that the buyer can get a similar property at a price lower than the asking price. Given: $B_1 = \{F, F \Rightarrow P, A \Rightarrow \neg P\}$ and $B_2 = \{F, F \Rightarrow P, A \Rightarrow \neg P\}$,*

and $PE \succ_1 P\bar{E} \succ_1 \bar{P}E \succ_1 \bar{P}\bar{E}$ and $\bar{P}\bar{E} \succ_2 P\bar{E} \succ_2 \bar{P}E \succ_2 PE$, a bargaining problem $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ can be defined.

REMARK: Our bargaining model defined above is quite similar to the ordinal bargaining model defined by Shubik [14]. In this model, a bargaining situation (between two players) can be represented as a tuple $(A, D, \geq_1, \geq_2)$, where A is the set of possible agreements, D is the disagreement (i.e., the outcome when the agents fail to reach an agreement), and $\geq_1$ and $\geq_2$ are the preference orderings over the set $A \cup \{D\}$ of the agents 1 and 2, respectively.[2] In the ordinal bargaining model, a bargaining solution F maps from every bargaining situation $(A, D, \geq_1, \geq_2)$ to an agreement $f(A, D, \geq_1, \geq_2) \in A$ (see e.g., [12] for more details).

In the following, we'll explore the cases when the agents' beliefs define their bottom-lines (and thus, the disagreement position of the agents), and also when the agents' beliefs are uncertain and can be revised (in relation to the opponents' beliefs and position).

3.1 Bargaining with fixed bottom lines

When the agents' beliefs are not revisable, they define the agents' disagreement points. We can now discuss a reformulation of Nash's axioms for the logical bargaining model by associating agent i's disagreement point with his beliefs. Firstly, we will define the disagreement point for a logical bargaining problem. Intuitively, the notion of disagreement point (or threat point) has been used to encode a bargainer's bargaining power. That is, the higher the utility of the disagreement point to a bargainer, the more power she has as she can walk away from the negotiation and obtain that high utility outside of the negotiation. This matches with our designation of the agent's beliefs. It's her beliefs that define what she thinks she and her opponent can get outside of the negotiation, which subsequently defines her relative bargaining power and the negotiation disagreement point.

DEFINITION 1. Let $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ be a bargaining problem, agent i's **disagreement point** is the outcome $D_i \in \mathcal{O}$ such that (i) D_i is least preferred to i, and (ii) D_i is consistent with B_i. Then, agent i's **bargaining power** is defined to be the number of outcomes ruled out by D_i, according to agent i's preference ordering:

$$U_i(D_i) = \#\{o \in \mathcal{O} : D_i \succ_i o\}.$$

An outcome $o \in \mathcal{O}$ is **agreement-feasible** if $o \succeq_i D_i$ for $i = 1, 2$.

For bargaining with ordinal preference, it has been shown by Osborne and Rubinstein [12] that a reformulation of **A1** (i.e., an axiom expressing Invariance of Equivalent Preference Representations) would result in unattractive bargaining solutions.

In the rest of this paper, we'll denote the agent other than agent i by $-i$. Furthermore, for convenience of presentation, given two outcomes $o, o' \in \mathcal{O}$, we'll say that $o' \succeq o$ if and only if $o' \succeq_i o$ for $i = 1, 2$; similarly, $o' \approx o$ if and only if $o' \succeq o$ and $o \succeq o'$.

Pareto Efficiency axiom can be formulated in our model as follows.

PE. If $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ is a bargaining problem with $o, o' \in \mathcal{O}$ and $o' \succeq o$ and $o' \succ_j o$ for some $j \in \{1, 2\}$, then $o \notin f(BP)$.

Note that in this paper, by Pareto-efficiency we mean the Strong Pareto Efficiency, rather than the Weak Pareto Efficiency which

---

[2]The preference orderings are required to be total pre-orders. That is, a complete transitive reflexive binary relation.

states that an outcome is only inefficient when there is some other outcome that can improves for *all* agents.

Next, the axiom of Independence of Irrelevant Alternatives will be formulated in our model as follows.

IIA. If $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ and $BP' = (\mathcal{O}', \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ are bargaining problems with $\mathcal{O} \subseteq \mathcal{O}'$, and $f(BP') \subseteq \mathcal{O}$ then $f(BP) = f(BP')$.

Finally, axiom Symmetry can be interpreted as imposing the requirement of fairness on a bargaining solution. That is, when the bargaining situation is symmetric for the two bargainers in the sense that they have similar bargaining power and that there is no possible agreement that can provide a particular payoff structure to the two bargainers without another possible agreement that can provide the opposite payoff structure, then the solution should give the bargainers the same payoffs. With a discrete set of alternatives, it can not be guaranteed to have an alternative that satisfies this property.

EXAMPLE 2. *Continue with our running example, consider the following bargaining situation:* $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ *where* $B_1 = \{E\}$ *and* $B_2 = \{E\}$ *(i.e., both bargainers insist in an early settlement), and* $PE \succ_1 P\bar{E} \succ_1 \bar{P}E \succ_1 \bar{P}\bar{E}$ *and* $\bar{P}E \succ_2 P\bar{E} \succ_2 \bar{P}\bar{E} \succ_2 P\bar{E}$. *It's easy to see that the two agents have similar bargaining power in which they would both rule out the two least preferred outcomes and it happens that, in this bargaining situation, they share the same set of outcomes they are willing to agree on, namely the set* $\{PE, \bar{P}E\}$. *However, they have opposite preferences over the outcomes in this set. Thus, neither outcome would be an attractive solution in this bargaining situation. By allowing this set of outcomes to be the solution in this situation, we are open to any resolution, including Nash's suggestion of non-physical agreements such as the lotteries over these outcomes.*

To ensure fairness, we will target outcomes that aim at maximising the payoffs for agents with the smallest gains (in utility). We define a cardinal gain of an outcome o for an agent i to be the number of outcomes between o and the disagreement point D_i. Formally,

DEFINITION 2. Let $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ be a bargaining problem and an agreement-feasible outcome $\sigma \in \mathcal{O}$, the (cardinal) **gain** of outcome σ for agent i is defined as follows:

$$G_i(\sigma) = \#\{o \in \mathcal{O} : \sigma \succ_i o \ \& \ o \succeq_i D_i\}.$$

Basically, the axiom for fairness, to be presented in the following, ensures that the difference between the bargainers' gains would be minimal. However, to avoid the bargainers to settle for fair but suboptimal outcomes, we require only that fairness be subject to the optimality of the outcome. We will introduce a concept to allow efficiency to be formulated in unanimity, namely *Unanimous Efficiency* (UE). Intuitively, if o' can improve for both the worst-off agent and the better-off agent in comparison to o then o is considered to be UE-dominated by o' and should not be selected as a bargaining agreement. Formally,

DEFINITION 3. Let $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ be a bargaining problem with the agreement-feasible outcomes $o, o' \in \mathcal{O}$, o is **UE-dominated** by o' if $\min_{i=1,2}(G_i(o')) \geq \min_{i=1,2}(G_i(o))$ and $\max_{i=1,2}(G_i(o')) \geq \max_{i=1,2}(G_i(o))$, and at least one of them has to be a strict inequality.

Moreover, we'll also say that two outcomes o and o' are **UE-equivalent** *if and only if* $\min_{i=1,2}(G_i(o')) = \min_{i=1,2}(G_i(o))$ *and* $\max_{i=1,2}(G_i(o')) = \max_{i=1,2}(G_i(o))$.

An outcome is **unanimously efficient** *if it is not UE-dominated by any other outcome.*[3]

LEMMA 1. *Let $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ be a bargaining problem with an agreement-feasible outcome $o \in \mathcal{O}$, if o is unanimously efficient then it is also Pareto efficient.*

Before proving Lemma 1, we state a convention to be used throughout the rest of the paper: Given an outcome o, if $G_1(o) = G_2(o)$, we assume that $\arg\min_{i=1,2}(G_i(o))$ will pick out a single value, either 1 or 2 (which one would be picked doesn't matter). Furthermore, if $j = \arg\min_{i=1,2}(G_i(o))$ then $\arg\max_{i=1,2}(G_i(o)) = -j$. That is, $\arg\min_{i=1,2}(G_i(o))$ (resp. $\arg\max_{i=1,2}(G_i(o))$) is guaranteed to deterministically return a single agent j (resp. $-j$) whose cardinal gain $G_j(o)$ (resp. $G_{-j}(o)$) is smallest (resp. largest).

Proof: Suppose, to the contrary, that o is not Pareto efficient. That is, there are $o' \in \mathcal{O}$ and $j \in \{1, 2\}$ such that $o' \succ_j o$ and $o' \succeq_{-j} o$. Obviously, o' is agreement-feasible. We'll consider two cases:

Case 1: $\arg\min_{i=1,2}(G_i(o')) = \arg\min_{i=1,2}(G_i(o)) = k$. If $k = j$ then $\min_{i=1,2}(G_i(o')) > \min_{i=1,2}(G_i(o))$ and $\max_{i=1,2}(G_i(o')) \geq \max_{i=1,2}(G_i(o))$. If $k \neq j$ then $\max_{i=1,2}(G_i(o')) > \max_{i=1,2}(G_i(o))$ and $\min_{i=1,2}(G_i(o')) \geq \min_{i=1,2}(G_i(o))$. Either way, we have o is UE-dominated by o', and thus can not be unanimously efficient.

Case 2: $\arg\min_{i=1,2}(G_i(o')) \neq \arg\min_{i=1,2}(G_i(o))$. Without loss of generality, we can assume that $\arg\min_{i=1,2}(G_i(o')) = 1$ and $\arg\min_{i=1,2}(G_i(o)) = 2$. That is, $G_1(o') = \min_{i=1,2}(G_i(o'))$ and $G_2(o) = \min_{i=1,2}(G_i(o))$ and $G_2(o') = \max_{i=1,2}(G_i(o'))$ and $G_1(o) = \max_{i=1,2}(G_i(o))$. Therefore, $G_1(o) \geq G_2(o)$ and $G_2(o') \geq G_2(o)$. Also, because $o' \succ_j o$ for some $j \in \{1, 2\}$, at least one of the above inequality has to be strict. Since $o' \succ_j o$ for some $j \in \{1, 2\}$ and $o' \succeq_{-j} o$,[4] $G_1(o') \geq G_1(o)$.

Hence, $G_1(o') \geq G_1(o) \geq G_2(o)$ and $G_2(o') \geq G_1(o') \geq G_1(o)$, and at least one of the inequalities $G_1(o) \geq G_2(o)$ and $G_2(o') \geq G_1(o')$ has to be strict. In other words, $\min_{i=1,2}(G_i(o')) \geq \min_{i=1,2}(G_i(o))$ and $\max_{i=1,2}(G_i(o')) \geq \max_{i=1,2}(G_i(o))$, and at least one of these inequalities has to be strict.

Therefore, o is UE-dominated by o', and thus can not be unanimously efficient. □

The following *Fairness* axiom can now be formulated.

FR. If $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ is a bargaining problem with an agreement-feasible outcome $o \in \mathcal{O}$. If there is an agreement-feasible and unanimously efficient outcome $o' \in \mathcal{O}$ such that $|G_1(o) - G_2(o)| > |G_1(o') - G_2(o')|$ then $o \notin f(BP)$.

In other words, axiom **FR** allows us to select the fairer outcomes among those that are unanimously efficient.

In addition to the above Fairness axiom, we will also require that when outcomes are UE-equivalent, the bargaining solution will not be biased towards a particular one. The following axiom formulates this requirement for unbiasedness.

UB. If $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ is a bargaining problem with agreement-feasible and UE-equivalent outcomes $o, o' \in \mathcal{O}$. Then, $o \in f(BP)$ if and only if $o' \in f(BP)$.

Finally, we will replace the Pareto Efficiency (**PE**) axiom by a stronger one, requiring that the bargaining solution be unanimously efficient, rather than only Pareto-efficient.

UE. If $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ is a bargaining problem with the agreement-feasible outcomes $o, o' \in \mathcal{O}$ and o is UE-dominated by o' then $o \notin f(BP)$.

Obviously, by Lemma 1, if the bargaining solution f satisfies **UE** then it also satisfies **PE**. In the following, we'll develop a bargaining solution that is uniquely identified by the axioms **UE**, **FR** and **UB**. The developed solution is based on the well-known *egalitarian* solution:

THEOREM 1. *There is a bargaining solution $f^E : \mathcal{BP} \rightarrow 2^{\mathcal{O}}$ given by:*

$$f^E(\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle) = \arg\max_{o \in ES}(\max_{i=1,2} G_i(o)), \text{ where}$$

$$ES = \arg\max_{o \in AF}(\min_{i=1,2} G_i(o))$$

*where $AF \subseteq \mathcal{O}$ denotes the set of agreement-feasible outcomes of BP.[5] Then, a bargaining solution $f : \mathcal{BP} \rightarrow 2^{\mathcal{O}}$ satisfies **UE**, **FR**, and **UB** if and only if $f = f^E$.*

Before proving Theorem 1, we'll discuss a relevant result. Note that, $\approx$ is an equivalence relation on the set $\mathcal{O}$. Moreover, if we'll only focus on the set $AF \subseteq \mathcal{O}$ of agreement-feasible outcomes of BP, then AF can be reduced to a collection of equivalence classes, such that each equivalence class can be represented by a pair $(G_1(o), G_2(o))$, where o is a member of the equivalence class.

LEMMA 2. *If $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ is a bargaining problem then $f^E(BP)$ contains at most two equivalence classes of the agreement-feasible outcomes of BP. Moreover, if $f^E(BP)$ contains exactly one equivalence class then it has to be of the form (g, g) (i.e., it gives both agents the same gain); if $f^E(BP)$ contains exactly two equivalence classes then they have to be of the form (g, h) and (h, g).*

Proof (of the Lemma): Obviously, from the definition of the function f^E, there exist $k_{min}, k_{max} \geq 0$ such that for every outcome $o \in ES$, $\min_{i=1,2} G_i(o) = k_{min}$ and for every outcome $\sigma \in f^E(BP)$, $\max_{i=1,2} G_i(\sigma) = k_{max}$. Thus, when $k_{min} = k_{max}$, $f^E(BP)$ contains exactly one equivalence class that is of the form (g, g), where $g = k_{min} = k_{max}$.

Otherwise, $k_{min} \neq k_{max}$ and $f^E(BP)$ contains exactly two equivalence classes of the form (g, h) and (h, g) where $g = k_{min}$ and $h = k_{max}$. □

Proof (of the Theorem): First we prove that, f^E satisfies **UE**, **FR**, and **UB**.

That f^E satisfies **UE**: Let $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ be a bargaining problem. Let $\sigma \in f^E(BP)$, we'll prove that σ is unanimously efficient. Suppose, to the contrary, that there is an agreement-feasible outcome $o \in AF$ such that $\min_{i=1,2} G_i(o) \geq \min_{i=1,2} G_i(\sigma)$ and $\max_{i=1,2} G_i(o) \geq \max_{i=1,2} G_i(\sigma)$ and at least one of these inequalities is strict. Moreover, there exists $k_{min} \geq 0$ such that $\forall a \in AF. \min_{i=1,2} G_i(a) \leq k_{min}$ and $k_{min} = \min_{i=1,2} G_i(\sigma)$. Thus, $k_{min} = \min_{i=1,2} G_i(o) = \min_{i=1,2} G_i(\sigma)$. In other words, $o \in ES$. However, according to Lemma 2, there exists $k_{max} \geq 0$ that, together with k_{min}, characterises the equivalence classes defining $f^E(BP)$ and $\forall a \in ES. \max_{i=1,2} G_i(a) \leq k_{max}$ and $k_{max} = \max_{i=1,2} G_i(\sigma)$. Thus, $\max_{i=1,2} G_i(a) \leq \max_{i=1,2} G_i(\sigma)$. Contradiction.

---

[3] Note that, similar to our remark about Pareto-efficiency, our definition of Unanimous Efficiency is also a strong one.

[4] We use $-j$ to denote the agent other than j.

[5] Note that more precise notations would be AF_{BP} and ES_{BP}. But since the bargaining problem BP is always clear from the context, we'll drop these subscripts.

That f^E satisfies **FR**: Suppose, to the contrary, that there is a bargaining problem $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ and $\sigma \in f^E(BP)$ such that there is a unanimously efficient outcome $o \in AF$ and $|G_1(\sigma) - G_2(\sigma)| > |G_1(o) - G_2(o)|$.

According to Lemma 2, σ belongs to the equivalence classes characterised by the two non-negative numbers k_{min} and k_{max} (possibly equal to each other).

We have $k_{min} = \min_{i=1,2} G_i(\sigma)$ and $k_{max} = \max_{i=1,2} G_i(\sigma)$.

Case 1: $\min_{i=1,2} G_i(o) = k_{min}$. Since $|G_1(\sigma) - G_2(\sigma)| > |G_1(o) - G_2(o)|$, clearly $\max_{i=1,2} G_i(o) < k_{max}$, which is a contradiction because o is UE-dominated by σ.

Case 2: $\min_{i=1,2} G_i(o) < k_{min}$. Since o is unanimously efficient, it is the case that $\max_{i=1,2} G_i(o) > k_{max}$. But then $|G_1(\sigma) - G_2(\sigma)| = k_{max} - k_{min} < \max_{i=1,2} G_i(o) - \min_{i=1,2} G_i(o) = |G_1(o) - G_2(o)|$. Contradiction.

It's obvious that f^E satisfies **UB**.

We are now proving that a solution f satisfying **UE**, **FR**, and **UB** necessarily obtains f^E.

Suppose, to the contrary that, there exists a solution f satisfying **UE**, **FR**, and **UB** and a bargaining problem $BP = (\mathcal{O}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ such that $f(BP) \neq f^E(BP)$. First, we'll show that it's not possible for $f(BP) \setminus f^E(BP) \neq \emptyset$. Assume by way of contradiction that there exists an outcome $o \in AF$ such that $o \in f(BP) \setminus f^E(BP)$. Then o is unanimously efficient and is not ruled out by axiom **FR**. According to Lemma 2, the set of outcomes $f^E(BP)$ can be partitioned into equivalence classes characterised by the non-negative numbers k_{min} and k_{max} (possibly equal to each other). Clearly, $\min_{i=1,2} G_i(o) \leq k_{min}$. If $\min_{i=1,2} G_i(o) = k_{min}$ then, for o to be unanimously efficient, $\max_{i=1,2} G_i(o) \geq k_{max}$. Thus, $\max_{i=1,2} G_i(o) = k_{max}$ In other words, $o \in f^E(BP)$. Contradiction. If $\min_{i=1,2} G_i(o) < k_{min}$ then, for o to be unanimously efficient, $\max_{i=1,2} G_i(o) > k_{max}$. But then, there is a unanimously efficient outcome $\sigma \in f^E(BP)$ such that $|G_1(\sigma) - G_2(\sigma)| = k_{max} - k_{min} < \max_{i=1,2} G_i(o) - \min_{i=1,2} G_i(o) = |G_1(o) - G_2(o)|$. Thus, $o \notin f(BP)$ according to axiom **FR**. Contradiction.

That $f^E(BP) \setminus f(BP) \neq \emptyset$ follows trivially from axiom **UB** and Lemma 2. $\square$

It's also straightforward to see that f^E satisfies axiom **IIA**.

COROLLARY 1. *The bargaining solution f^E defined in Theorem 1 satisfies* **IIA**.

So far in this section, the agent's beliefs B_i are only used to define the agent's disagreement point and don't play much role in characterising the negotiation outcome. The problem becomes more challenging when we allow the agents' beliefs to change according to the bargaining situation.

3.2 Bargaining with revisable agents' beliefs

According to the bargaining model introduced in the preceding section, when the agents' beliefs B_i define the bottom-lines D_i that result in an empty set of agreement-feasible outcomes AF (i.e., $\{o \in \mathcal{O} : o \succeq_1 D_1\} \cap \{o \in \mathcal{O} : o \succeq_2 D_2\} = \emptyset$), agreement is not possible. Nevertheless, in most negotiations, the agents' beliefs represent their inclination toward a particular position rather than unmovable. For instance, a buyer of a house may know for certain that an identical house was sold a month ago for \$y, and thus is not too willing to pay much more than \$y for this house. However, knowing that there is no other house left in the area that he can buy, he is perhaps willing to pay more than \$y, if there are compelling reasons for him to do so (e.g., there are other buyers who would like to buy a house in the area). In this situation, if P denotes the proposition that the vendor agent's asking price is higher than \$y,

then $\neg P$ doesn't necessarily define the buyer's disagreement point. It could be the case that the buyer believes in $\neg P$, but is also willing to retract this belief when learning about the scarcity of houses as well as the high demand for houses in the area.

Given that the agents' beliefs (and constraints) will play a crucial role in this model of bargaining, we will make the agents' hard constraints explicit in our bargaining model. In the rest of the paper, we will assume that the set of feasible negotiation outcomes is defined by the hard constraints $\mathcal{C}$. We denote by $\mathcal{O}_{\mathcal{C}}$ the set of feasible outcomes that satisfy the hard constraints $\mathcal{C}$. That is, $o \in \mathcal{O}_{\mathcal{C}}$ if and only if $[\![o]\!] \cap [\![\mathcal{C}]\!] \neq \emptyset$.

Given the propositional language $\mathcal{L}$ and the non-empty alphabet $\mathcal{P}_O$ for the negotiation outcomes, a bargaining problem is defined to be a tuple $BP = (\mathcal{C}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$, where $\mathcal{C} \subseteq \mathcal{L}$ is the set of hard constraints shared by all agents, $B_i \subseteq \mathcal{L}$ is the set of agent i's beliefs, and $\succeq_i$ is agent i's preference relation over the set $\mathcal{O}_{\mathcal{C}}$. Subsequently, the (movable) disagreement points $D_i \in \mathcal{O}_{\mathcal{C}}$ are defined such that D_i is least preferred to agent i and D_i is consistent with B_i. Moreover, the set of agreement-feasible outcomes AF_{BP} is defined to be $\{o \in \mathcal{O}_{\mathcal{C}} : o \succeq_1 D_1\} \cap \{o \in \mathcal{O}_{\mathcal{C}} : o \succeq_2 D_2\}$.[6] We will first state a trivial lemma:

LEMMA 3. *Let $BP = (\mathcal{C}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ be a bargaining problem, if the set $\mathcal{C} \cup B_1 \cup B_2$ is consistent then $AF_{BP} \neq \emptyset$.*

Proof: Let ω be a model of $\mathcal{C} \cup B_1 \cup B_2$. Let $o \in \mathcal{PO}$ be such that $\omega \in [\![o]\!]$. Clearly, $o \in \mathcal{O}_{\mathcal{C}}$ and o is consistent with both B_1 and B_2. Thus, $o \succeq_1 D_1$ and $o \succeq_2 D_2$. Thus, $o \in AF_{BP}$. $\square$

On the other hand, when the set $\mathcal{C} \cup B_1 \cup B_2$ is inconsistent, it doesn't always mean that $AF_{BP} = \emptyset$.

EXAMPLE 3. *Continue with our running example and consider the following bargaining situation $BP = (\mathcal{C}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$, where $\mathcal{C} = \{P\}$ (e.g., the agent receives the instruction from the vendor not to sell the house for less than the asking price and this is common knowledge), and $B_1 = \{E\}$ and $B_1 = \{\neg E\}$. Also, suppose that $P\bar{E} \succ_1 PE$ and $PE \succ_2 P\bar{E}$. Clearly, $\mathcal{C} \cup B_1 \cup B_2$ is inconsistent, but $D_1 = PE$ and $D_2 = P\bar{E}$, and $AF_{BP} = \{PE, P\bar{E}\}$.*

Given Lemma 3, one fairly naive idea is to perform belief merging (see e.g., [7]) with integrity constraint (i.e., merging B_1 and B_2 with the integrity constraint $\mathcal{C}$) to obtain a consistent belief base X which will be treated as the shared bottom line for both agents 1 and 2. Subsequently, the bargaining model described in Section 3.1 can be applied to characterise the negotiation outcome. Unfortunately, this straightforward idea will not work, for the simple reason that belief merging takes into account the two belief bases B_1 and B_2 when merging them with respect to the integrity constraints $\mathcal{C}$. But it fails to take into consideration the agents' preferences $\succeq_1$ and $\succeq_2$ regarding the preferred outcomes.

EXAMPLE 4. *In the running example, consider a bargaining situation in which $\mathcal{C} = \{\top\}$ (i.e., no hard constraints), $B_1 = \{P, \neg E\}$, and $B_2 = \{E\}$. Furthermore, the agents preferences are: $PE \succ_1 P\bar{E} \succ_1 \bar{P}E \succ_1 \bar{P}\bar{E}$ and $\bar{P}E \succ_2 \bar{P}\bar{E} \succ_2 PE \succ_2 P\bar{E}$ with $D_1 = P\bar{E}$ and $D_2 = PE$. Clearly, $\mathcal{C} \cup B_1 \cup B_2$ is inconsistent and most standard belief merging mechanisms (see [7]) would result in a merge belief base $X = \Delta_{\mathcal{C}}(B_1 \sqcup B_2) = \{P\}$. By taking X to define the common bottom line for both agents, the new disagreement points for them become $D_1' = D_1$ and $D_2' = P\bar{E}$.*

---

[6] When clear from the context, we will omit the subscript and write AF instead.

Clearly, this has disadvantaged agent 2. Moreover, it has also imposed agent 1's bottom line regarding attribute P on agent 2 without any reasonable justification and compensation.

On the other hand, any mechanism that searches for a negotiation outcome based only on the agents' preferences $\succeq_1$ and $\succeq_2$ without taking into account the agents' beliefs is likely to produce impractical outcomes as well. For instance, in the bargaining situation discussed at the beginning of this section, assume that $PE \succ_1 P\bar{E} \succ_1 \bar{P}E \succ_1 \bar{P}\bar{E}$ and $\bar{P}\bar{E} \succ_2 \bar{P}E \succ_2 P\bar{E} \succ_2 PE$. Assume also that the vendor's asking price is $\$x > \$y$, and $B_1 = \{P, \neg E\}$ (i.e., the vendor agent knows that the house he is selling is currently the only house for sale in the area and there are several buyers who are looking for a house in the area, while a recent government regulation requires mortgage lender to carry out a number of checks before releasing the fund for settlement), and $B_2 = \{\neg P, \neg E\}$ (i.e., the buyer knows that an identical house was sold for $\$y$ last month). As the set of agreement-feasible outcomes AF is empty, the agents need to make concessions to possibly reach an agreement. Without taking into account the agents' beliefs, the bargaining strategy of minimal concession (see [5]) suggests the following negotiation process: First, agent 2, the buyer, will make a minimal concession from its current offer of $\bar{P}\bar{E}$ to $\bar{P}E$; then, agent 1 makes a minimal concession, and accept the offer $\bar{P}E$. This outcome is certainly not justified since the right price in this case should be $\$x$ (thus, agreeing on P) while the agents can also agree on $\neg E$.

Therefore we argue that a reasonable mechanism should enable the agents to use their beliefs to make the decision on what concession to make, taking into account their preferences. Consequently, we'll investigate a strategic model for bargaining in the following section.

4. AN ARGUMENTATION-BASED MODEL OF BARGAINING

The axiomatic bargaining model introduced above is inherently static in the sense that only the outcome, and not the bargaining process, is analysed. This ensures a number of advantages such as tractability. Nevertheless, in most circumstances, it's important to study the bargaining process as well as the bargainers' strategies. For instance, we may be interested in knowing how the bargaining outcome is affected by changes in the bargaining procedure, and what would be the best strategy or decision a bargainer should take in a given situation.

The bargaining protocol to be used by the agents to reach an agreement is based on the belief negotiation models proposed by Booth [3]. In this model, the negotiation proceeds in rounds. The negotiation starts off with the initial offer profile $\vec{\Theta}^0 = (\Theta_1^0, \Theta_2^0)$, where an offer Θ_i^j is a subset of the set of feasible outcomes $\mathcal{O}$, indicating the outcomes agent i is willing to accept after j rounds of negotiation. If $\vec{\Theta}^j$ ($j \geq 0$) is consistent then the set of agreement-feasible outcomes $AF = \Theta_1^j \cap \Theta_2^j$ is non-empty and a physical agreement can be selected from AF. If $\vec{\Theta}^j$ ($j \geq 0$) is inconsistent then a "contest" between the agents will be carried out to select a subset of agents who are required to "make some concessions".[7] The new offer profile $\vec{\Theta}^{j+1}$ after the selected agents making the concession allows the negotiation to proceed to the next round. Under "certain predefined conditions", a disagreement is reached. Furthermore, under a monotonic concession protocol (see [13]), the new offer profile $\vec{\Theta}^{j+1}$ is required to include the previous one; i.e.,

---

[7]The generality of this protocol allows it to encompass many common negotiation protocols including the alternating-offer protocol and the simultaneous-concession protocol.

$\vec{\Theta}^j \subseteq \vec{\Theta}^{j+1}$), where $\vec{S} = (S_1, \ldots, S_n) \subseteq (T_1, \ldots, T_n) = \vec{T}$ if and only if $S_i \subseteq T_i$ for all $i \in \{1, \ldots, n\}$. A more precise bargaining protocol will be introduced later in this section.

From the discussion in the preceding section, a bargaining outcome should be based on the agents' beliefs about the bargaining situation at hand while taking into account their preferences. To formulate the idea that an agent's beliefs that define her position on certain bargaining issues can be undermined or dominated by her opponent's beliefs, we appeal to the argumentation-based framework [4, 1]. In a strategic model of bargaining, a bargaining problem $BP = (\mathcal{C}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ is based on a common language $\mathcal{L}$ and a common outcome alphabet $\mathcal{P}_O$, with the set of hard constraints $\mathcal{C}$ being common knowledge while the agents beliefs and preferences $\langle B_i, \succeq_i \rangle$ for $i = 1, 2$ are their private information. We will first reproduce some notions of argumentation theory.

DEFINITION 4. ([1]) *An* **argument** *of a set of sentences* $X \subseteq \mathcal{L}$ *(aka.* X-**argument**) *under the constraints* $\mathcal{C}$ *is a pair* (H, h), *where* $h \in \mathcal{L}$ *and* $H \subseteq X$, *satisfying:*

(i) $\mathcal{C} \cup H$ *is consistent,*

(ii) $\mathcal{C} \cup H \models h$, *and*

(iii) H *is minimal (i.e., no strict subset of* H *satisfies* (i) *and* (ii)*).*

H *is called the* **support** *and* h *the* **conclusion** *of the argument* (H, h). *Moreover, given two arguments* (H, h) *and* (H', h'), *if* $H \Leftrightarrow H'$ *and* $h \Leftrightarrow h'$ *then we treat them as the same argument. That is, a set of arguments can not contain both arguments.*

An argument (H', h') *is a* **subargument** *of the argument* (H, h) *iff* $H' \subseteq H$.

Given a set of arguments Γ, *the* **base** *of* Γ *is the set:* $\mathcal{B}_\Gamma = \bigcup_{(H,h) \in \Gamma} H$.

We denote by $\mathbb{A}_\mathcal{C}(X)$ the set of all X-arguments under the constraints $\mathcal{C}$.

DEFINITION 5. ([1]) *Let* (H, h) *and* (H', h') *be two arguments of* $\mathbb{A}_\mathcal{C}(X)$:

- (H, h) **rebuts** (H', h') *if and only if* $h \Leftrightarrow \neg h'$.

- (H, h) **undercuts** (H', h') *if and only if* $h \Leftrightarrow \neg h''$ *for some* $h'' \in H'$.

When (H, h) *rebuts or undercuts* (H', h'), *we also say that* (H, h) **attacks** (H', h'). *When* (H, h) *attacks* (H', h') *and* (H'', h'') *attacks* (H, h), *we say that* (H'', h'') **defends** (H', h').

We are now in a position to formally define our bargaining protocol. First, we define the notion of bargaining proposal.

DEFINITION 6. *Let* $BP = (\mathcal{C}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ *be a bargaining problem and* $\mathcal{O}_\mathcal{C}$ *denote the set of feasible outcomes. A* **bargaining proposal** *(or,* **proposal***) by agent* i *at stage* j, *denoted by* ρ_i^j *is a pair* (Θ_i^j, Γ_i^j), *where* $\Theta_i^j \subseteq \mathcal{O}_\mathcal{C}$ *is the set of outcomes agent* i *is willing to agree on and* $\Gamma_i^j \subseteq \mathbb{A}_\mathcal{C}(\mathcal{L})$ *is the set of arguments agent* i *has used to defend her offers* Θ_i^j.

The pair (ρ_1^j, ρ_2^j) *of the agents' proposals in stage* j *is called the* **bargaining context** *at stage* j.

It's important to note that, in a strategic model of bargaining the agents beliefs and preferences $\langle B_i, \succeq_i \rangle$ are their private information. Therefore, agent i can introduce arguments that are not based on her beliefs if she thinks that they would give her an advantage. In the following, we'll write $\bigvee \Theta$ instead of $\bigvee_{o \in \Theta} o$. We will now define the notion of an argument being relevant to a bargaining context.

DEFINITION 7. *Let (ρ_1, ρ_2) be a bargaining context, where $\rho_i = (\Theta_i, \Gamma_i)$ is agent i's proposal, for $i = 1, 2$. An argument (H, h) is* **relevant** *to this bargaining context (for agent i) if $(H, h) \notin \Gamma_i$ and*

- $\bigvee \Theta_{-i} \wedge h \models \bot$; *or*

- (H, h) *attacks an argument $(H', h') \in \Gamma_{-i}$.*

Of the two non-trivial conditions above, the former says that agent i rejects agent $-i$'s current offered outcomes Θ_{-i} by advancing an argument (H, h) that contradicts Θ_{-i} and thus requires agent $-i$ to make a concession. The latter allows agent i to advance an argument (H, h) to defeat a relevant argument (H', h') advanced by agent $-i$ in previous rounds of bargaining.

In the bargaining protocol informally described at the beginning of this section, for the "contest" to select who need to make a revised proposal during a negotiation round, we assume that all agents will have to submit the updated proposal in each round. Furthermore, an agent i's proposal in round 0 has to contain a non-empty offer $\Theta_i^0 \neq \emptyset$ and, to simplify the protocol, it also contains an empty set of arguments $\Gamma_i^0 = \emptyset$. Agent i's proposal in round $j > 0$, $\rho_i^j = (\Theta_i^j, \Gamma_i^j)$ is required to meet the following conditions:

- $\Theta_i^j \supseteq \Theta_i^{j-1}$;

- $\Gamma_i^j \supseteq \Gamma_i^{j-1}$ such that $\mathcal{B}_{\Gamma_i^j} \cup \mathcal{C}$ is consistent and, if $\Gamma_i^j \neq \Gamma_i^{j-1}$ then the new arguments have to be relevant to the previous bargaining context $(\rho_1^{j-1}, \rho_2^{j-1})$.

If $\vec{\Theta}^j$ $(j \geq 0)$ is consistent then the set of agreement-feasible outcomes $AF = \Theta_1^j \cap \Theta_2^j$ is non-empty and an agreement can be selected from AF. If $\vec{\Theta}^j$ $(j \geq 0)$ is inconsistent then the bargaining proceeds to the next round. If in two consecutive rounds of bargaining, the bargaining context is not updated, i.e., for some $j \geq 0, \rho_i^j = \rho_i^{j+1} = \rho_i^{j+2}$ for $i = 1, 2$, then the bargaining reaches a disagreement.

LEMMA 4. *The bargaining protocol defined above terminates.*

Proof: Since the alphabet $\mathcal{P}$ of the language $\mathcal{L}$ (and the alphabet $\mathcal{P}_O \subseteq \mathcal{P}$ of the bargaining outcomes) is finite, there can only a finite number of logically different arguments; i.e., the set $\mathbb{A}_{\mathcal{C}}(\mathcal{L})$ is finite and the set $\mathcal{O}_{\mathcal{C}}$ is also finite. Thus, if the bargaining does not terminate with a disagreement then at some point, both agents will have exhausted the set of arguments and thus will have to increasingly add the members of $\mathcal{O}_{\mathcal{C}}$ in their respective offers and the bargaining terminates with a non-empty set of agreement-feasible outcomes AF. □

Given the negotiation protocol, our example in the preceding section about the vendor agent who argues to convince the buyer to change her position on the price of the house can be described as follows.

EXAMPLE 5. *We will assume that the alphabet $\mathcal{P}$ also contains the following propositional symbols: Y for "a similar house was sold for \$y last month", M for "house prices last month reflects today market", S for "houses in the area have become scarce", D for "there has been an increase in the demand for houses in the area", C for "the market has changed with the price on the up", and R for "bargainers should exercise reciprocity". We'll also assume that $PE \succ_1 P\bar{E} \succ_1 \bar{P}E \succ_1 \bar{P}\bar{E}$ and $\bar{P}E \succ_2 \bar{P}\bar{E} \succ_2 P\bar{E} \succ_2 PE$ are the agents' respective preferences. That is, the buyer's preference on the value of E (whether it should be an early settlement) is conditional on the value of P (whether she has to pay the higher price).*

The negotiation starts off with the initial proposal profile $((\{PE\}, \emptyset), (\{\bar{P}E\}, \emptyset))$. In the next round, the buyer introduces the argument $(\{Y, M, Y \wedge M \Rightarrow \neg P\}, \neg P)$ and the seller introduces the argument $(\{Y, S, D, S \wedge D \Rightarrow C, Y \wedge C \Rightarrow P\}, P)$ to defend their respective positions. In the consecutive round, the seller then defeat the buyer's argument with the argument $(\{C, C \Rightarrow \neg M\}, \neg M)$. This settles the issue on the price of the house with the buyer making a concession and willing to accept any outcome from the set $\{\bar{P}E, \bar{P}\bar{E}, P\bar{E}\}$. However, since the set of agreement-feasible outcome $AF = \{PE\} \cap \{\bar{P}E, \bar{P}\bar{E}, P\bar{E}\}$ remains empty, the buyer then advances the argument $(\{P, R, P \wedge R \Rightarrow \neg E\}, \neg E)$. Since the seller has no counter-argument, he makes a concession and is willing to accept any outcome from the set $\{PE, P\bar{E}\}$. They settle with the outcome $P\bar{E}$.

To formalise the notion of winning argument in an exchange between bargainers, we'll appeal to the argument-based semantics of admissibility, introduced by Dung [4].

DEFINITION 8. *A set Γ of arguments is* **conflict-free** *if there are no two arguments (H, h) and (H', h') such that (H, h) attacks (H', h') or (H', h') attacks (H, h).*

A set Γ of arguments is **admissible** *if it is conflict-free and it defends all of its members against all attackers.*

A set Γ of arguments is **strongly admissible** *if it is admissible and none of the arguments that attack its members belong to an admissible set of arguments.*

The following lemma is trivial since at all stages j of the bargaining, $\mathcal{B}_{\Gamma_i^j} \cup \mathcal{C}$ is required to be consistent.

LEMMA 5. *The sets of arguments contained in the bargaining proposals introduced by the agents according to the bargaining protocol defined above are conflict-free.*

The following axiom requires that bargainers do not ignore strongly admissible sets of arguments that support an agent's position.

SA. If $BP = (\mathcal{C}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ is a bargaining problem and $AG \subseteq \mathcal{O}_{\mathcal{C}}$ the agreement reached after j rounds of bargaining. If the set of arguments $\Gamma \subseteq \Gamma_i^j$ is strongly admissible then for each outcome $o \in AG$: $\bigwedge_{(H,h) \in \Gamma} h \wedge o$ is consistent.

Intuitively, axiom **SA** requires that, if agent i can present an argument that agent $-i$ can not defeat, then every agreed outcome has to be consistent with the conclusions obtainable from this set of arguments.

We can now state a lemma trivially derived from the definition of strongly admissible sets of arguments.

LEMMA 6. *Given a bargaining problem BP, if a sentence h is supported by a strongly admissible set of arguments Γ from the current bargaining context (ρ_1^j, ρ_2^j) then there does not exist any admissible set of arguments Γ' from the current bargaining context that supports $\neg h$.*

Equipped with Lemma 6 and assuming a belief revision operator $*_{AGM}$ that satisfies the AGM axioms (see [6]), we can now define an argument-augmented bargaining problem of a given bargaining problem.

DEFINITION 9. *Let $BP = (\mathcal{C}, \langle B_1, \succeq_1 \rangle, \langle B_2, \succeq_2 \rangle)$ be a bargaining problem and given a a belief revision operator $*_{AGM}$ that satisfies the AGM axioms. Consider the set of all admissible sets of*

arguments of the base $B_1 \cup B_2$: $AS_{BP} = \{\Gamma \subseteq \mathbb{A}_{\mathcal{C}}(B_1 \cup B_2) : \Gamma$ is strongly admissible $\}$. Let α denote $\bigwedge_{(H,h) \in \bigcup_{\Gamma \in AS_{BP}} \Gamma} h$, then the **argument-augmented bargaining problem** *of BP, denoted by $\mathcal{A}^{BP}$ is defined to be $(\mathcal{C}, \langle B_1^{*AGM}\alpha, \succeq_1 \rangle, \langle B_2^{*AGM}\alpha, \succeq_2 \rangle)$.*

Given a bargaining problem BP, the following bargaining solution can be defined:

$$f^{A\text{-}E}(BP) = \begin{cases} \text{disagreement} & \text{if } AF_{\mathcal{A}^{BP}} = \emptyset \\ f^E(\mathcal{A}^{BP}) & \text{otherwise} \end{cases}$$

The following theorem is obvious (given Lemma 6 and the AGM axioms):

THEOREM 2. *The bargaining solution $f^{A\text{-}E}$ satisfies axiom* **SA**.

Given the negotiation protocol above and our proposed argument-based bargaining framework, we are interested in finding the equilibrium strategies in a bargaining situation. Note that, in a strategic model of bargaining, a bargainer's beliefs and preferences are her private information and the bargaining progresses when the agents exchange their proposals, in our bargaining protocol, by simultaneously putting them on the negotiation table. However, in the presence of incomplete information, the agents clearly have the incentives not to reveal their true preferences and beliefs. Therefore, to develop a tractable strategic model of bargaining, we'll need to make a number of assumptions including an enforceable penalty mechanism which is also underlying the negotiation framework developed by Rosenschein and Zlotkin [13, 17]. Given such assumptions, we have developed a symmetric Nash equilibrium strategy for the bargainers based on the minimal concession strategy introduced by Dung et al. [5]. This consideration is beyond the scope of the present paper and will be included in our future work.

5. CONCLUSION AND FUTURE WORK

In this paper we introduced an axiomatic model of bargaining with logical beliefs and goals for the purpose of bargaining analysis. To the best of our knowledge, our model is the first logic-based axiomatic model of bargaining that does not suffer the problem of syntax-sensitivity while still ensuring that our proposed bargaining solution is uniquely characterized by a concise set of intuitive axioms (see e.g., [16, 15]). This is the most important contribution of our paper. Moreover, our framework allows for a separation between the bargainers' beliefs and their respective goals. This is important because not all beliefs of the agents that are relevant to the negotiation will necessarily end up on the negotiation table. Many of them may be used only for the agents to make decision about whether to accept an offer or what counter-offer to be made.

We have also taken into account the problem of dynamic negotiation in which the bargainers' bottom lines could be changed during the negotiation. The problem is challenging, particularly in the context of incomplete information. We appeal to the formalism of argumentation framework to allow for the accommodation of new and revised beliefs and their effects on the bargainers' bottom lines. We are currently investigating a strategic model of bargaining to complement our axiomatic model. This is also the final step in realising the famous Nash program. This work will be reported in our future papers.

From a multi-agent systems point of view, it is always an important question to investigate the computational complexity of the solutions and the concepts we have proposed. This also remains a challenge to be addressed in the future work.

6. REFERENCES

[1] L. Amgoud and C. Cayrol. Inferring from inconsistency in preference-based argumentation frameworks. *J. Autom. Reason.*, 29(2):125–169, 2002.

[2] K. Binmore. *Game Theory and the Social Contract, Vol. 2: Just Playing*. The MIT Press, Mar. 1998.

[3] R. Booth. Social contraction and belief negotiation. *Information Fusion*, 7(1):19–34, 2006.

[4] P. M. Dung. On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming and n-person games. *Artificial Intelligence Journal*, 77:321–357, 1995.

[5] P. M. Dung, P. M. Thang, and F. Toni. Towards argumentation-based contract negotiation. In *Procs. of Computational Models of Argument: COMMA*, pages 134–146, 2008.

[6] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. The MIT Press, 1988.

[7] S. Konieczny, J. Lang, and P. Marquis. DA2 merging operators. *Artificial Intelligence*, 157(1-2):49 – 79, 2004.

[8] T. Meyer, N. Foo, R. Kwok, and D. Zhang. Logical foundations of negotiation: outcome, concession and adaptation. In *Proceedings of the 19th National Conference on Artificial Intelligence (AAAI-04)*, pages 293–298, 2004.

[9] T. Meyer, N. Foo, R. Kwok, and D. Zhang. Logical foundations of negotiation: strategies and preferences. In *Proceedings of the 9th International Conference on the Principles of Knowledge Representation and Reasoning (KR'04)*, pages 311–318, 2004.

[10] J. Nash. The bargaining problem. *Econometrica*, 18(2):155–162, 1950.

[11] J. Nash. Two-person cooperative games. *Econometrica*, 21(1):129–140, 1953.

[12] M. J. Osborne and A. Rubinstein. *Bargaining and Markets*. Academic Press, 1990.

[13] J. S. Rosenschein and G. Zlotkin. *Rules of Encounter: Designing Conventions for Automated Negotiation Among Computers*. MIT Press, Cambridge, Massachusetts, 1994.

[14] M. Shubik. *Game Theory in the Social Sciences: Concepts and Solutions*. MIT Press, Cambridge, 1982.

[15] D. Zhang. A logic-based axiomatic model of bargaining. *Artif. Intell.*, 174(16-17):1307–1322, 2010.

[16] D. Zhang and Y. Zhang. An ordinal bargaining solution with fixed-point property. *J. Artif. Int. Res.*, 33:433–464, November 2008.

[17] G. Zlotkin and J. S. Rosenschein. Mechanisms for automated negotiation in state oriented domains. *Journal of Artificial Intelligence Research*, 5:163–238, 1996.