

# A Formal Analysis of Interest-based Negotiation\*

Iyad Rahwan<sup>1,2</sup>

<sup>1</sup>*Faculty of Informatics, The British University in Dubai  
P.O.Box 502216, Dubai, UAE (irahwan@acm.org)*

<sup>2</sup>*School of Informatics, University of Edinburgh, UK*

Philippe Pasquier

*School of Interactive Art and Technology  
Simon Fraser University, Vancouver, Canada*

Liz Sonenberg

*Dept. of Information Systems, University of Melbourne  
Parkville, VIC 3010 Australia*

Frank Dignum

*Dept. of Information & Computing Sciences  
Utrecht University, Utrecht, The Netherlands*

June 14, 2009

**Keywords:** Agents, multiagent systems, negotiation, argumentation

## Abstract

In multi-agent systems (MAS), negotiation provides a powerful metaphor for automating the allocation and reallocation of resources. Methods for automated negotiation in MAS include auction-based protocols and alternating offer bargaining protocols. Recently, argumentation-based negotiation has been accepted as a promising alternative to such approaches. Interest-based negotiation (IBN) is a form of argumentation-based negotiation in which agents exchange (1) information about their underlying goals; and (2) alternative ways to achieve these goals. However, the usefulness of IBN has been mostly established in the literature by appeal to intuition or by use of specific examples. In this paper, we propose a new formal model for reasoning about interest-based negotiation protocols. We demonstrate the usefulness of this framework by defining and analysing two different IBN protocols. In particular, we characterise conditions that guarantee their advantage (in the sense of expanding the set of individual rational deals) over the more classic proposal-based approaches to negotiation.

---

\*This is a revised and expanded version of a paper that appeared in the proceedings of AAAI 2007 [31].

# 1 Introduction

Negotiation is the search for agreement on the exchange (or allocation) of scarce resources among (self-)interested parties. Approaches to one-to-one<sup>1</sup> automated negotiation have been classified in three categories [20]: (1) game theoretic (2) heuristic and (3) argumentation based.

The first two families are characterized by the exchange of offers between parties with conflicting positions and are commonly referred to as *proposal-based* approaches. That is, agents exchange proposed agreements—in the form of bids or offers—and when proposed deals are not accepted, the possible response is either a counter-proposal or withdrawal. Argumentation-based negotiation (ABN) approaches, on the other hand, enable agents to exchange additional *meta*-information (i.e. arguments) during negotiation [32]. This paper is concerned with a particular style of argument-based negotiation, namely *interest-based negotiation* (IBN) [33], a form of ABN in which agents explore and discuss their underlying interests. Information about other agents’ goals may be used in a variety of ways, such as discovering and exploiting common goals.

Most existing literature supports the claim that ABN is useful by presenting specific examples that show how ABN can lead to agreement where a more basic exchange of proposals cannot (e.g. the mirror/picture example in [25]). The focus is usually on underlying semantics of arguments and argument acceptability. However, no formal analysis exists of how agent preferences, and the range of possible negotiation outcomes, change as a result of exchanging arguments.

Our aim here is to explore how exchanging meta-information about the agent’s underlying goals can help improve the negotiation process. To this end, we explore situations where agents generate their preferences using a deliberation procedure that results in hierarchies of goals. This abstraction is common and has been used in the context of automated planning [13] and multi-agent coordination [8]. We also abstract away from the underlying argumentation logic. We use this simplified framework to characterise precisely how agent preferences and the set of possible negotiation outcomes change as a result of exchanging information about agents’ goals. To our knowledge, this constitutes the first formal analysis of the outcomes of interest-based negotiation, and how they may differ from proposal-based approaches, namely alternating-offer bargaining. We then present two simple IBN protocols. The first one (IBNP1) allows agents to reveal their underlying goals motivating the negotiation when asked. We show that under certain conditions (e.g. that agents’ goals do not interfere with each other), revealing underlying goals always leads to an expansion of the set of possible deals. The second protocol (IBNP2) extends the first one by allowing agents to reveal private knowledge that can be used to generate alternative—and previously unknown—ways to achieve the goal revealed. Here again, we show that using IBN, agents can only increase the utility of a given offer using this interest-based negotiation strategy. As such, our new framework begins bridging the gap between the theory and practice of ABN, and our analysis provides a step towards understanding the dynamics of more complex IBN and ABN dialogues.

---

<sup>1</sup>Many-to-many and many-to-one automated negotiations are usually handled using a growing variety of auction-based models [39, 38] and these negotiation types are not considered in this work.

This paper makes three key contributions to the state-of-the-art in automated negotiation. Firstly, the paper provides the first framework for systematically analysing interest-based negotiation protocols. This framework enables the analysis of goal-related arguments in negotiation while abstracting away from the underlying logical framework for argumentation. This makes the framework applicable to studying issues arising in a variety of specific instantiations of ABN (including [25]).

The second main contribution of this paper is in providing the first systematic formal analysis of the outcomes of two key interest-based negotiation protocols: one that allows goal revelation, and one that allows the exchange of previously unknown partial plans. In particular, the paper characterises general conditions under which these protocols are guaranteed to expand the set of possible deals, compared to traditional offer-based, alternating-offer protocols.

Thirdly, by providing a formal framework and demonstrating its usefulness in analysing IBN protocols, this paper begins bridging the gap between the theory and practice of IBN. In particular, understanding how negotiation outcomes change as a result of interest-based arguments is crucial to revealing the dynamics of IBN protocols. Indeed, the dynamics of IBN dialogues are much less understood than those of bargaining protocols.

After presenting preliminaries in the next Section, we present a basic bargaining protocol in Section 3. Then, in Section 4, we provide a framework for capturing agents' underlying interests. In Section 5 we discuss how agents' knowledge of each others' interests influences their preferences positively. We use the framework to analyse some key properties of two IBN protocols in Sections 6 and 7. We present a discussion in Section 9 and summarise related work in Section 8 before concluding the paper in Section 10.

## 2 Preliminaries

Our negotiation framework consists of a set of two *agents*  $\mathcal{A}$  and a finite set of *resources*  $\mathcal{R}$ , which are indivisible and non-sharable. An *allocation of resources* is a partitioning of  $\mathcal{R}$  among agents in  $\mathcal{A}$  [12].

**Definition 1. (Allocation)** An allocation of resources  $\mathcal{R}$  to a set of agents  $\mathcal{A}$  is a function  $\Lambda : \mathcal{A} \rightarrow 2^{\mathcal{R}}$  such that  $\Lambda(i) \cap \Lambda(j) = \{\}$  for  $i \neq j$  and  $\bigcup_{i \in \mathcal{A}} \Lambda(i) = \mathcal{R}$

Agents may have different preferences over sets of resources, defined in the form of utility functions. At this stage, we do not make any assumptions about the properties of preferences/utility functions (e.g. being additive, monotonic, etc.).

**Definition 2. (Utility functions)** Every agent  $i \in \mathcal{A}$  has a utility function  $u_i : 2^{\mathcal{R}} \rightarrow \mathbb{R}$ .

Given their preferences, agents may be able to benefit from reallocating (i.e. exchanging) resources. Such reallocation is referred to as a *deal*. A rational self-interested agent should not accept deals that result in loss of utility. However, we will make use of *side payments* in order to enable agents to compensate each other for accepting deals that result in loss of utility [12].

**Definition 3. (Payment)** A payment is a function  $p : \mathcal{A} \rightarrow \mathbb{R}$  such that  $\sum_{i \in \mathcal{A}} p(i) = 0$ ,

Note that the definition ensures that the total amount of money is constant. If  $p(i) > 0$ , the agent *pays* the amount  $p(i)$ , while  $p(i) < 0$  means the agent *receives* the amount  $-p(i)$ . We can now define the notion of ‘deal’ formally.

**Definition 4. (Deal)** Let  $\Lambda$  be the current resource allocation. A deal with money is a tuple  $\delta = (\Lambda, \Lambda', p)$  where  $\Lambda'$  is the suggested allocation,  $\Lambda' \neq \Lambda$ , and  $p$  is a payment.

Let  $\Delta$  be the set of all possible deals. By overloading the notion of utility and the symbol  $u_i$ , we will also refer to the utility of a deal (as opposed to the utility of an allocation) defined as follows.

**Definition 5. (Utility of a Deal for an Agent)** The utility of deal  $\delta = (\Lambda, \Lambda', p)$  for agent  $i$  is:

$$u_i(\delta) = u_i(\Lambda'(i)) - u_i(\Lambda(i)) - p(i)$$

A deal is *rational* for an agent only if it results in positive utility for that agent, since otherwise, the agent would prefer to stick with its initial resources.

**Definition 6. (Rational Deals for an Agent)** A deal  $\delta$  is rational for agent  $i$  if and only if  $u_i(\delta) > 0$

If a deal is rational for each individual agent given some payment function  $p$ , it is called *individual rational*.

**Definition 7. (Individual Rational Deals)** A deal  $\delta$  is individual rational if and only if  $\forall i \in \mathcal{A}$  we have  $u_i(\delta) \geq 0$  and  $\exists j \in \mathcal{A}$  such that  $u_j(\delta) > 0$ .

In other words, no agent becomes worse off, while at least one agent becomes better off. This is equivalent to saying that the new allocation *Pareto dominates* the initial allocation, given the payment. We denote by  $\Delta^* \subseteq \Delta$  the *set of individual rational deals*.

### 3 Bargaining Protocol

An *offer* (or *proposal*) is a deal presented by one agent which, if accepted by the other agents, would result in a new allocation of resources. In the alternating-offer (or bargaining) protocol, agents exchange proposals until one is found acceptable or negotiation terminates (e.g. because a deadline was reached or the set of all possible proposals were exhausted without agreement). In this paper, we will restrict our analysis to two agents  $i$  and  $j$  with  $i \neq j$ . The bargaining protocol initiated by agent  $i$  with agent  $j$  is shown in Table 1.

Bargaining can be seen as a search through possible allocations of resources. In the brute force method, agents would have to exchange every possible offer before a deal is reached or disagreement is acknowledged. The number of possible allocations of resources to agents is  $|\mathcal{A}|^{|\mathcal{R}|}$ , which is exponential in the number of resources. The number of possible offers is even larger, since agents would have to consider not only

**Bargaining Protocol 1 (BP1):**

Agents start with resource allocation  $\Lambda^0$  at time  $t = 0$ . At each time  $t > 0$ :

1.  $\text{propose}(i, \delta^t)$ : Agent  $i$  proposes to  $j$  deal  $\delta^t = (\Lambda^0, \Lambda^t, p^t)$  not proposed earlier;
2. Agent  $j$  either:
  - (a)  $\text{accept}(j, \delta^t)$ : accepts, and negotiation terminates with deal  $\delta^t$ ; or
  - (b)  $\text{reject}(j, \delta^t)$ : rejects, and negotiation terminates with no deal; or
  - (c) counter-proposes by going to step 1 at time  $t + 1$  with agents' roles swapped.

Table 1: Basic bargaining protocol

every possible allocation of resources, but also every possible payment.<sup>2</sup> Various computational frameworks for bargaining have been proposed in order to enable agents to reach deals quickly. For example, Faratin et al [14] use a heuristic for generating counter proposals that are as similar as possible to the previous offer they rejected.

We characterise the set of deals that are *reachable* using *any* given protocol. The set of reachable deals can be conveniently characterised in terms of the history of offers made (thus, omitting, for now, other details of the protocol). To enable studying changes in the utility function later in the paper, we will superscript utility functions with time-stamps.

**Definition 8. (Dialogue History)** A dialogue history of protocol  $P$  between agents  $i$  and  $j$  is an ordered sequence  $h$  of tuples consisting of a proposal and a utility function (over allocations) for each agent

$$h = \langle (\delta^1, u_i^1, u_j^1), \dots, (\delta^n, u_i^n, u_j^n) \rangle$$

where  $t = 1, \dots, n$  represents time.

**Definition 9. (Protocol-Reachable Deal)** Let  $P$  be a protocol. A deal  $\delta^t$  is  $P$ -reachable if and only if there exists two agents  $i$  and  $j$  which can generate a dialogue history according to  $P$  such that  $\delta^t$  is offered by some agent at time  $t$  and  $\delta^t$  is individual rational given  $u_i^t, u_j^t$ .

Note that whether or not agents will actually reach a particular deal under a given protocol depends not only on the protocol, but also on the strategies employed by agents.

## 4 Underlying Interests

In most existing alternating-offer bargaining negotiation frameworks, agents' utility functions over possible deals are assumed to be *pre-determined* (e.g. as weighted sums)

<sup>2</sup>Since payments are real numbers, to guarantee termination, the protocol should enforce a time limit, or a limit on the range of possible payments/concession made.

and *fixed* throughout the interaction. That is, throughout the dialogue history,  $u_i^1 = \dots = u_i^n$  for any agent  $i$ .

We now present a framework for capturing the interdependencies between goals at different levels of abstraction. Although this framework is simpler than those in the planning literature, its level of abstraction is sufficient for our purpose.

Let  $\mathcal{G} = \{g_1, \dots, g_m\}$  be a finite set of all possible goals. And let  $sub : \mathcal{G} \times 2^{\mathcal{G} \cup \mathcal{R}}$  be a relationship between a goal and the sub-goals or resources needed to achieve it.

Intuitively,  $sub(g, \{g_1, \dots, g_n\})$  means that achieving all the goals  $g_1, \dots, g_n$  results in achieving the higher-level goal  $g$ . Each sub-goal in the set  $\{g_1, \dots, g_n\}$  may itself be achievable using another set of sub-goals, thus resulting in a goal hierarchy. We assume that this hierarchy takes the form of a tree (called *goal tree* or *plan*), in which each goal appears only once. This implies that no goal contributes to one of its sub-goals. This condition is reasonable since the sub-goal relation captures specialisation of abstract goals into more concrete goals.

Note that  $sub$  is a relation, not a function, to allow us to express goals that have multiple sets of *alternative* sub-goals/resources. Hence, there may be multiple possible plans for achieving a goal.

**Definition 10. (Partial plan)** A partial plan for achieving goal  $g_0$  is a tree  $T$  such that:

- $g_0$  is the root;
- Each non-leaf node is a goal  $g \in \mathcal{G}$  with children  $x_1, \dots, x_n \in \mathcal{G} \cup \mathcal{R}$  such that  $sub(g, \{x_1, \dots, x_n\})$ ; i.e. among alternatives for achieving  $g$ , only one is selected.
- Each leaf node is  $x_i \in (\mathcal{R} \cup \mathcal{G})$ ;

A complete plan is a goal tree in which all leaf nodes are resources.

**Definition 11. (Complete plan)** A complete plan for achieving goal  $g_0$  is a partial plan  $T$  in which each leaf node  $r_i \in \mathcal{R}$ .

**Example 1.** Suppose we have a set of goals  $\mathcal{G} = \{g_1, \dots, g_4\}$  and a set of resources  $\mathcal{R} = \{r_1, \dots, r_6\}$  such that  $sub(g_1, \{g_2, g_3\})$ ,  $sub(g_1, \{g_2, g_4\})$ ,  $sub(g_2, \{r_1, r_2\})$ ,  $sub(g_3, \{r_3, r_4\})$ ,  $sub(g_4, \{r_5, r_6\})$ . Suppose also that the agent's main goal is  $g_1$ . Figure 1 shows three plans that can be generated. Tree  $T_1$  is a partial plan (since goal  $g_3$  is a leaf node), while  $T_2$  and  $T_3$  are (the only possible) complete plans for achieving  $g_1$ .

Let  $gnodes(T) \subseteq \mathcal{G}$  be the set of goal nodes in tree  $T$ . And let  $leaves(T) \subseteq \mathcal{R} \cup \mathcal{G}$  be the set of leaf nodes in tree  $T$ . Let  $rleaves(T) = leaves(T) \cap \mathcal{R}$  be the set of resource leaves. And similarly, let  $gleaves(T) = leaves(T) \cap \mathcal{G}$  be the set of goal leaves. Note that for a complete plan  $T$ ,  $leaves(T) = rleaves(T)$ , that is, leaf nodes contain resources only.

Let  $\mathcal{T}$  be the set of all (partial or complete) plans that can be generated in the system, and let  $\mathcal{T}(g)$  be the set of all plans that have  $g$  as a root.

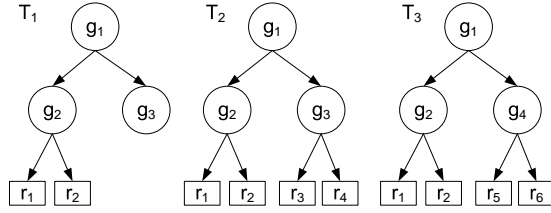


Figure 1: Partial plans ( $T_1$ ) and complete plans ( $T_2, T_3$ )

**Definition 12. (Individual Capability)**

An agent  $i \in \mathcal{A}$  with resources  $\Lambda(i)$  is individually capable of achieving goal  $g \in \mathcal{G}$  if and only if there is a complete plan  $T \in \mathcal{T}(g)$  such that  $leaves(T) \subseteq \Lambda(i)$

We assume that each agent  $i$  is assigned a single goal  $G(i) \in \mathcal{G}$  that it needs to achieve, and we refer to it as the agent’s *main goal*.<sup>3</sup> We further assume that agent  $i$  assigns a worth to this goal  $worth_i(G(i)) \in \mathbb{R}$ .

**Example 2.** Following on Example 1, suppose agent  $i$  with goal  $G(i) = g_1$  has resources  $\Lambda(i) = \{r_1, r_2, r_3, r_4, r_5\}$ . Agent  $i$  is individually capable of achieving  $g_1$  through complete plan  $T_2$ , since  $leaves(T_2) \subseteq \Lambda(i)$ .

Note that the agent also has the option of retaining its resources and not using them to achieve its goal (e.g. they are worth more than the goal). Here, we say that the agent has selected the *null plan*, denoted  $\check{T}$ . We can characterise the set of all complete plans that an agent can choose from.

**Definition 13. (Individually Achievable Plans)** The set of plans that can be achieved by agent  $i$  individually using allocation  $\Lambda(i)$  is:

$$\mathcal{T}_{\Lambda(i)} = \{T \in \mathcal{T} : leaves(T) \subseteq \Lambda(i)\} \cup \{\check{T}\}$$

We now want to provide a new definition of the utility of an allocation, which takes into account the agent’s underlying goal. Therefore, we differentiate between the *intrinsic* value of the resource and its potential contribution to a goal. So, if the agent’s resources cannot be used to achieve its goals, then the utility of these resources will be the sum of their intrinsic values, as above. If, on the other hand, the agent is able to achieve its goal using some of its resources, then the utility calculation must take into account the difference between the utility gained by achieving the goal and the utility lost by consuming the resources.

The agent must select the *best* plan, i.e. the plan that minimizes the cost of the resources used. To capture this, let  $v_i : \mathcal{R} \rightarrow \mathbb{R}$  be a valuation function such that  $v_i(r)$  is agent  $i$ ’s private valuation of resource  $r$ . Then we can define the cost incurred by agent  $i$  in executing plan  $T$  as:  $cost_i(T) = \sum_{r \in leaves(T)} v_i(r)$ . Then, we can define the *utility of a plan* as follows.<sup>4</sup>

<sup>3</sup>Multiple goals can be expressed by a single goal that has one possible decomposition.

<sup>4</sup>Note that so far, we have different notions of utility: the utility of an allocation, the utility of a plan, and the utility of a deal.

**Definition 14. (Utility of a Plan)** Let  $i$  be an agent with goal  $G(i)$ . And let  $\mathcal{T}_i^*$  be the set of available alternative plans  $i$  can choose from. The utility of plan  $T \in \mathcal{T}_i^*$  for agent  $i$  is a function  $\tilde{u}_i : \mathcal{T}_i^* \rightarrow \mathbb{R}$  is defined as follows:

$$\tilde{u}_i(T) = \begin{cases} 0 & \text{if } T = \check{T}, \\ \text{worth}_i(G(i)) - \text{cost}_i(T) & \text{otherwise} \end{cases}$$

Note that for agent  $i$  with allocation  $\Lambda(i)$  and goal  $G(i)$ , the set of available alternatives (not considering other agents in the system) is  $\mathcal{T}_i^* = (\mathcal{T}_{\Lambda(i)} \cap \mathcal{T}(G(i)))$ .

Since the null plan does not achieve a goal and does not incur any cost, the agent retains all its initial resources, and therefore the utility of the null plan is zero.

**Example 3.** Following on Example 1, suppose agent  $i$  with goal  $G(i) = g_1$  has resources  $\Lambda(i) = \{r_1, r_2, r_3, r_4, r_5, r_6\}$ . Suppose also that  $\text{worth}_i(g_1) = 85$  and resource valuations  $v_i(r_1) = 20$ ,  $v_i(r_2) = 10$ ,  $v_i(r_3) = 6$ ,  $v_i(r_4) = 5$ ,  $v_i(r_5) = 8$ ,  $v_i(r_6) = 7$ . Then, we have:

$$\begin{aligned} \tilde{u}_i(T_2) &= 85 - (20 + 10 + 6 + 5) = 44 \\ \tilde{u}_i(T_3) &= 85 - (20 + 10 + 8 + 7) = 40 \\ \tilde{u}_i(\check{T}) &= 0 \end{aligned}$$

We now define the utility of an allocation for an agent. Note that this is a specialisation of the general utility function in Definition 2. Note also that underlying our framework is the assumption that resources are consumable, at least for the period in question, in the sense that a single resource cannot be used simultaneously in multiple plans. An example of a consumable resource is “fuel” consumed to run an engine, or “server time” consumed to run a scientific experiment.

**Definition 15. (Utility)** Let  $i \in \mathcal{A}$  be an agent with goal  $G(i)$ . The utility of set of resources  $R \subseteq \mathcal{R}$  is defined using a function  $u_i : 2^{\mathcal{R}} \rightarrow \mathbb{R}$  such that:

$$u_i(R) = \max_{T \in \mathcal{T}_i^*} \tilde{u}_i(T)$$

The utility of a deal remains defined as above.

**Example 4.** Following Example 3, the utility of the resources is  $u_i(\Lambda(i)) = 44$ , and the best plan is  $T_2$ .

## 5 Mutual Interests

One of the main premises of IBN is that agents may benefit from exploring each other’s underlying interests. For example, agents may avoid making irrelevant offers given each others’ goals. Knowledge of *common*<sup>5</sup> goals may help agents reach better agreements, since they may discover that they can benefit from goals achieved by one another. In this paper, we focus on the case of common goals.

<sup>5</sup>Note that common goals are different from individual goals of the same kind. Two agents may both want to hang the same picture, or may each want to hang a different picture.

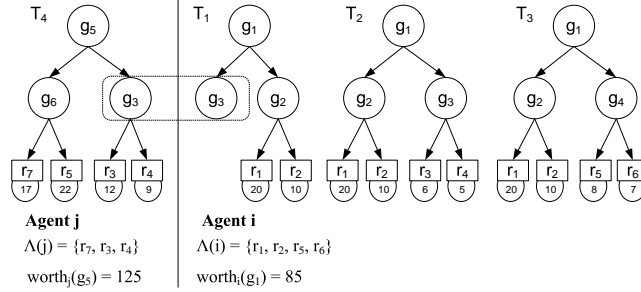


Figure 2: Agent  $i$  can benefit from  $j$ 's committed goal

We first formalise the idea that an agent may benefit from a goal (or sub-goal) achieved by another. Suppose an agent  $j$  is committed to some plan  $T_j$ , written  $I_j(T_j)$ . Then, another agent  $i$ , with  $I_i(T_i)$ , may benefit from the goals in  $gnodes(T_j)$  if one or more of these goals is part of  $T_i$ . Note, however, that not every goal in  $gnodes(T_j)$  is useful to  $i$ , but rather those goals for which  $j$  has a complete goal (sub-)tree. Thus, we define the notion of *committed goals*.

**Definition 16. (Committed Goals)** Let  $i \in \mathcal{A}$  be an agent with resources  $\Lambda(i)$  with  $I_i(T_i)$  at time  $t$ . The committed goals of  $i$  at time  $t$  is denoted  $cgoals_i^t$  and defined as:  $cgoals_i^t = \{g \in gnodes(T_i) : g \text{ has a plan } T \in \mathcal{T}_{\Lambda(i)} \text{ where } T \text{ is a sub-tree of } T_i\}$

When there is no ambiguity, we shall drop the superscript  $t$  that denotes time.

**Definition 17. (Achievable Plans)** The set of partial plans that can be achieved by agent  $i$  using allocation  $\Lambda(i)$  given agent  $j$ 's committed goals  $cgoals_j^t$  at time  $t$  is:

$$\mathcal{T}_{\Lambda(i), cgoals_j^t} = \{T \in \mathcal{T} : \text{leaves}(T) \subseteq \Lambda(i) \cup cgoals_j^t\} \cup \checkmark$$

**Example 5.** Figure 2 shows agent  $i$  and  $j$  with goals  $g_1$  and  $g_5$  respectively, with all possible plans, the resources owned by every agents and, under every resource, the agent's private valuation. Note that  $T_2$  is possible but not achievable by  $i$  with  $\Lambda(i)$ . Now, suppose  $j$  intends plan  $T_4$ . This means that  $g_3 \in cgoals_j$ . While  $T_1$  is not individually-achievable, it is now a viable alternative for agent  $i$  to achieve  $g_1$  since agent  $j$  is committed to goal  $g_3$ .

The following lemma follows immediately.

**Lemma 1.** At any time  $t$ ,  $\mathcal{T}_{\Lambda(i)} \subseteq \mathcal{T}_{\Lambda(i), cgoals_j^t}$

*Proof.* Let  $T \in \mathcal{T}_{\Lambda(i)}$ . By definition 13,  $\text{leaves}(T) \subseteq \Lambda(i)$ , from which it follows that  $\text{leaves}(T) \subseteq \Lambda(i) \cup cgoals_j^t$ . By definition 17, we have  $T \in \mathcal{T}_{\Lambda(i), cgoals_j^t}$ .  $\square$

From the lemma, it follows that when agents take into account goals committed by other agents, the set of available plans expands, since agents are no longer restricted to considering complete plans. Formally, for agent  $i$  with goal  $G(i)$  and resources

**IBN Protocol 1 (IBNP1):**

Agents start with resource allocation  $\Lambda^0$  at time  $t = 0$ . At each time  $t > 0$

1.  $\text{propose}(i, \delta^t)$ : Agent  $i$  proposes to  $j$  deal  $\delta^t = (\Lambda^0, \Lambda^t, p^t)$  not proposed earlier;
2. Agent  $j$  either:
  - (a)  $\text{accept}(j, \delta^t)$ : accepts, and negotiation terminates with deal  $\delta^t$ ; or
  - (b)  $\text{reject}(j, \delta^t)$ : rejects, and negotiation terminates with no deal; or
  - (c) counter-proposes by going to step 1 at time  $t + 1$  with agents' roles swapped; or
  - (d) switches to IBN on  $\delta^t$ . Let  $dgoals_i^t = \emptyset$  for all  $i \in \mathcal{A}$  be agents' declared goals.
    - i.  $\text{why}(j, x)$ :  $j$  asks  $i$  for underlying goal for  $x \in \Lambda^t(i) \cup dgoals_i^t$ ;
    - ii.  $i$  either:
      - A.  $\text{assert}(i, I_i(g))$ :  $i$  states a goal, which is added to  $dgoals(i)$ ; or
      - B.  $\text{decline}(i)$ : declines giving the information;
    - iii.  $j$  either:
      - A.  $\text{accept}(j, \delta^t)$ :  $j$  accepts  $\delta^t$ , if now more favourable; or
      - B. seeks more information by going to step 2.d.i; or
      - C.  $\text{pass}(j)$ : the protocol moves to step 1 with agents' roles swapped.

Table 2: A simple IBN protocol

$\Lambda(i)$ , the set of available options at time  $t$  is now  $\mathcal{T}_i^* = (\mathcal{T}_{\Lambda(i), cgoals_j^t} \cap \mathcal{T}(G(i)))$ . Agents can now consider partial plans, as long as the missing parts of these plans are committed  $j$ . From this, it also follows that the utility of an allocation may increase. The example below calculates agent  $i$ 's utility for partial plan  $T_1$ , which was previously not considered.

**Example 6.** *Continuing on Example 5 and Figure 2. We now have  $\tilde{u}_i(T_1) = 85 - (20 + 10) = 55$ ,  $\tilde{u}_i(T_3) = 40$  and  $\tilde{u}_i(\bar{T}) = 0$  (recall that  $T_2 \notin \mathcal{T}_i^*$  for now). Therefore,  $u_i(\Lambda(i)) = 55$ . This contrasts with the calculation that does not take  $j$ 's goal into account, in which case  $u_i(\Lambda(i)) = 40$ .*

## 6 Case Study 1: Discovering Common Goals

We showed how agents' utilities of allocations may increase if agents have knowledge of each other's underlying goals. However, full awareness of other agents' goals is rarely achievable, especially when agents are self-interested. Agents may progressively (and selectively) reveal information about their goals using a variety of interaction protocols. For example, agents could reveal their entire goal trees at once, or may do so in a specific order. Moreover, agents may reveal their underlying goals symmetrically (e.g. simultaneously) or asymmetrically, etc. We now look at a specific IBN protocol and analyse it using the above concepts.

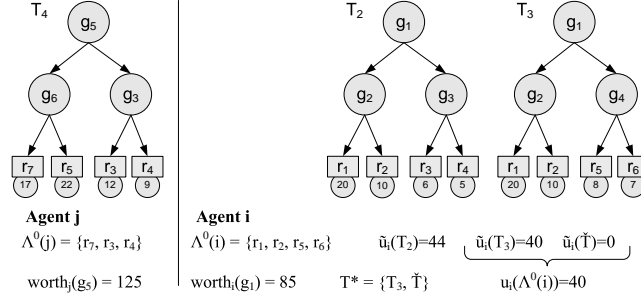


Figure 3: Initial stage of the IBN dialogue

We assume that agents have no prior knowledge of each other’s main goals or preferences; and that prior to negotiation, each agent  $i$  considers all individually-achievable plans, for its main goal, using  $\Lambda(i)$ , as well as potential rational deals. An IBN protocol is presented Table 2. Note that this protocol is asymmetric, since during the IBN sub-dialogue, the agent being questioned is assumed to *fix* its intended plans, while the questioning agent may accept the deal in question by discovering new viable plans that take into account the questionee’s goals.

Let us now consider an extension of the previous example.

**Example 7.** Suppose agent  $i$ ’s initial situation is as described in Figure 3. Here,  $i$  begins with two achievable plans:  $T_3$  and  $\tilde{T}$ . As shown in Example 6,  $u_i(\Lambda^0(i)) = 40$ . Suppose  $i$  considers acquiring resources  $\{r_3, r_4\}$  to enable possible plan  $T_2$ . With  $\{r_3, r_4\}$ ,  $\tilde{u}_i(T_2) = 85 - (20 + 10 + 6 + 5) = 44$ , so  $i$  would be willing to pay up to  $44 - 40 = 4$  units for  $\{r_3, r_4\}$ , since he would still be better-off than working solo. Agent  $j$  on the other hand only has one possible plan, which is  $T_4$  with utility  $\tilde{u}_j(T_4) = 125 - 60 = 65$ , but is unable to execute it because it needs  $r_5$ . Now, agent  $i$  initiates negotiation with  $j$ . The following is a possible sequence of proposals, in which  $i$  offers to buy  $r_3$  and  $r_4$  for payment 3, and  $j$  counter-offers to buy  $r_5$  for payment 9:<sup>6</sup>

1. propose( $i, (\Lambda^0, \Lambda^1, p^1)$ ), where  $\Lambda^1(i) = \{r_1, r_2, r_3, r_4, r_5, r_6\}$ ,  $\Lambda^1(j) = \{r_7\}$ ,  $p^1(i) = 3$ ,  $p^1(j) = -3$
2. propose( $j, (\Lambda^0, \Lambda^2, p^2)$ ), where  $\Lambda^2(i) = \{r_1, r_2, r_6\}$ ,  $\Lambda^2(j) = \{r_3, r_4, r_5, r_7\}$ ,  $p^2(i) = 9$ ,  $p^2(j) = -9$

At this point, agent  $i$  may attempt to know why  $j$  needs one of the resources it requested, say  $r_3$ , and the following follows:

4. why( $i, r_3$ )
5. assert( $j, I_i(g_3)$ )

<sup>6</sup>This is an arbitrary sequence of offers, since in this paper we are not concerned with the bargaining strategy that dictates the sequence of offers in the bargaining part of the protocol. Other techniques from the literature may be used for this purpose, e.g. [14].

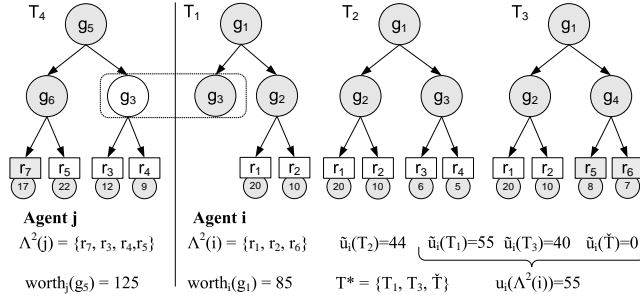


Figure 4: Stage 2 of the IBN dialogue

At this point, we have the situation in Figure 4 (the non-shaded parts represent revealed information). Plan  $T_1$  now becomes a viable option for  $i$ , which leads  $i$  to abandon its demand for  $r_3$  and  $r_4$ . Moreover, recall that  $\tilde{u}_i(T_1) = 55$ , so  $i$  can now give up resource  $r_5$  for payment 9 in a deal.

8.  $\text{accept}(i, (\Lambda^0, \Lambda^2, p^2))$

In summary,  $i$  gives up  $r_5$  in exchange for getting  $g_3$  and a payment of 5. While  $j$  pays 5 for  $r_5$  and achieves its goal (which was not possible before). Both agents gain utility, and the utilities of the deal  $\delta$  are as follows:

$$u_i(\delta) = u_i(\Lambda^2(i)) - u_i(\Lambda^0(i)) - p^2(i) = (55 - 8) - 40 + 9 = 16$$

$$u_j(\delta) = u_j(\Lambda^2(j)) - u_j(\Lambda^0(j)) - p^2(j) = 65 - 0 - 9 = 56$$

Note that in calculating the utility of  $i$ 's new allocation, we subtracted 8 since  $i$  has given up  $r_5$  in the deal, which it values as 8.

Let us now analyse IBNP1. We first show that IBNP1 subsumes BP1.

**Proposition 1.** *Every bargaining-reachable deal is also IBN-reachable.*

*Proof.* If in IBNP1, no agent ever switches to an interest-based dialogues –step (d), then the two algorithms BP1 and IBNP1 become identical. Hence, any deal reachable through bargaining is also reachable through IBN.  $\square$

We are mainly interested in how agents' perceptions of the utility of allocations change over time. Let  $dgoals : \mathcal{A} \rightarrow 2^{\mathcal{G}}$  be a function that returns the set of goals declared by an agent. We assume that agents do not lie about their goals, in the sense that they do not declare goals they are not committed to. Formally,  $dgoals_i^t \subseteq cgoals_i^t$  for any agent  $i$  at any given time  $t$ . Let  $\mathcal{T}_{\Lambda^t(i), dgoals_j^t} \subseteq \mathcal{T}$  be the set of goal trees that can be achieved by agent  $i$  using allocation  $\Lambda^t(i)$  given  $j$ 's declared goals  $dgoals_j^t$ , i.e.

$$\mathcal{T}_{\Lambda^t(i), dgoals_j^t} = \{T \in \mathcal{T} : \text{leaves}(T) \subseteq \Lambda^t(i) \cup dgoals_j^t\}$$

The below proposition then follows. The proposition shows that by using protocol IBNP1, the set of available plans for the inquiring agent expands, but never goes beyond

the set of plans that take into account all of the counterpart's actual goals. Formally, for agent  $i$  with goal  $G(i)$  and resources  $\Lambda(i)$ , the set of available options at time  $t$  is now  $\mathcal{T}_i^* = \mathcal{T}_{\Lambda(i), dgoals_j^t} \cap \mathcal{T}(G(i))$ . In other words, this result shows that the protocol is *sound* in the sense that it does not lead agents to produce incorrect plans.

**Proposition 2.** *At any time  $t$ ,  $\mathcal{T}_{\Lambda^t(i)} \subseteq \mathcal{T}_{\Lambda^t(i), dgoals_j^t} \subseteq \mathcal{T}_{\Lambda^t(i), cgoals_j^t}$*

*Proof.* Proof of  $\mathcal{T}_{\Lambda^t(i)} \subseteq \mathcal{T}_{\Lambda^t(i), dgoals_j^t}$  is similar to proof of Lemma 1. The fact that  $\mathcal{T}_{\Lambda^t(i), dgoals_j^t} \subseteq \mathcal{T}_{\Lambda^t(i), cgoals_j^t}$  follows from the assumption that  $dgoals_j^t \subseteq cgoals_j^t$ .  $\square$

The below proposition shows that the protocol is capable of providing an agent with all relevant information about the other agent's plan (provided the other agent is cooperative). In other words, the protocol is *complete* in the sense that it enables an agent to take into account all possible positive side effects from the counterpart's plan (by traversing the latter's entire goal tree).

**Proposition 3.** *Using the protocol IBNP1, at any time  $t$ , it is possible for any agent  $j$  to obtain complete knowledge of the entire goal structure of the intended plan by the other agent  $i$ , provided  $i$  does not decline to answer questions.*

*Proof.* At any given round  $t$ , suppose agent  $i$  intends arbitrary complete plan  $T_i^t \in \mathcal{T}$ , and proposes  $\delta^t$  (Step 1). By definition,  $leaves(T_i^t) \subseteq \Lambda^t(i)$ , i.e.  $i$  must obtain through  $\delta^t$  every resource needed for achieving  $T_i^t$ . After this request (Step 2.d),  $j$  could ask  $why(j, r)$  for each  $r \in leaves(T_i^t)$ . This would be done over  $|leaves(T_i^t)|$  iterations of Step 2.d. As a result,  $dgoals_j^t$  will contain the set of goals that are immediate parents of resource  $r \in leaves(T_i^t)$ . Similarly, Step 2.d could be repeated to obtain the immediate parents of those goals, until the main goal is revealed. Thus, every intended goal of  $i$  will eventually be in  $dgoals_j^t$ .  $\square$

The following proposition states that as the negotiation counterpart declares more of its goals, the inquirer's utility of any plan may increase, but can never decline. This is because the inquirer is increasingly able to account for the positive *side effects* of other agents' goals.

**Proposition 4.** *At any given time  $t$ , if the protocol is in stage 2.d initiated by agent  $i$ , as the set  $dgoals_j^t$  increases, the utility  $u_i(\delta^t)$  of the current proposal may only increase.*

*Proof.* Recall that the set of available alternative plans  $i$  can choose from is  $\mathcal{T}_i^* = \mathcal{T}_{\Lambda(i), dgoals_j^t} \cap \mathcal{T}(G(i))$ , and that  $\mathcal{T}_{\Lambda^t(i)} \subseteq \mathcal{T}_{\Lambda^t(i), dgoals_j^t}$ . It follows that as  $dgoals_j^t$  increases, the set  $\mathcal{T}_i^*$  also grows monotonically. Recall that  $u_i(\Lambda^t(i)) = \max_{T \in \mathcal{T}_i^*} \tilde{u}_i(T)$ . Hence, as  $u_i(\Lambda^t(i))$  is applied to maximise over a monotonically increasing set, its value can increase but not decrease. Consequently,  $u_i(\delta^t)$  is non-decreasing.  $\square$

It follows that at any time  $t$  where agent  $j$  intends plan  $T_j^t$  and  $i$  is inquiring  $j$ 's goals, as  $dgoals_j^t$  converges towards  $cgoals_j^t$ , then  $u_i(\Lambda^t(i))$  will reach the *objective* utility, that is the utility that reflects the true utility of  $\Lambda^t(i)$ .

Now we are ready to present the proposition below, which shows that IBN can lead to more deals than bargaining.

**IBN Protocol 2 (IBNP2):**

Agents start with resource allocation  $\Lambda^0$  at time  $t = 0$ . At each time  $t > 0$

1.  $\text{propose}(i, \delta^t)$ : Agent  $i$  proposes to  $j$  deal  $\delta^t = (\Lambda^0, \Lambda^t, p^t)$  not proposed earlier;
2. Agent  $j$  either:
  - (a)  $\text{accept}(j, \delta^t)$ : accepts, and negotiation terminates with deal  $\delta^t$ ; or
  - (b)  $\text{reject}(j, \delta^t)$ : rejects, and negotiation terminates with no deal; or
  - (c) counter-proposes by going to step 1 at time  $t + 1$  with agents' roles swapped; or
  - (d) switches to IBN on  $\delta^t$ . Let  $dgoals_i^t = \emptyset$  for all  $i \in \mathcal{A}$  be agents' declared goals.
    - i.  $\text{why}(j, x)$ :  $j$  asks  $i$  for underlying goal for  $x \in \Lambda^t(i) \cup dgoals_i^t$ ;
    - ii.  $i$  either:
      - A.  $\text{assert}(i, I_i(g))$ :  $i$  states a goal, which is added to  $dgoals(i)$ ; or
      - B.  $\text{decline}(i)$ : declines giving the information;
    - iii.  $j$  either:
      - A.  $\text{accept}(j, \delta^t)$ :  $j$  accepts  $\delta^t$ , if now more favourable; or
      - B. seeks more information by going to step 2.d.i; or
      - C.  $\text{achieves}(j, \text{sub}(g, \{x_1, \dots, x_n\}))$ ;  $j$  states an alternative to achieve  $g$  for some  $g \in dgoals_i^t$ .
      - D.  $\text{pass}(j)$ : the protocol moves to step 1 with agents' roles swapped.

Table 3: An IBN protocol that enables revealing new alternative goal decompositions

**Proposition 5.** *There may exist IBN-reachable deals that are not bargaining-reachable.*

*Proof.* Recall that a deal  $\delta^t$  is  $P$ -reachable under protocol  $P$  if and only if there exists two agents  $i$  and  $j$  which can generate a dialogue history according to  $P$  such that  $\delta^t$  is offered by some agent at time  $t$  and  $\delta^t$  is individual rational given  $u_i^t, u_j^t$ . Let  $\delta^t$  be a specific deal offered by  $i$  but that is not individual rational for  $j$  given  $u_j^t$ . However, from proposition 4, we know that declaring additional goals in a given round can only increase the utility of a given deal. Thus  $u_j^t(\delta^t)$  may increase, possibly making  $\delta^t$  individual rational for  $j$  (i.e. acceptable).  $\square$

From propositions 1 and 5 above, the following corollary follows.

**Corollary 1.** *The set of IBN-reachable deals is a super-set of bargaining-reachable deals.*

## 7 Case Study 2: Revealing New Plans

In the previous section, we analysed a protocol in which agents exchange information about their goals. But we assumed that both agents have complete knowledge of the

goal decomposition relation  $sub()$ . In other words, each agent knows every possible way of achieving its goals.

We now consider the case in which an agent may not be aware of some alternative plans of achieving some (sub-)goals. Exchanging this information may enable agents to reach agreements not previously possible. This was shown through the well-known painting/mirror hanging example presented by Parsons et al [25]. The example concerns two home-improvement agents – agent  $i$  trying to hang a painting, and agent  $j$  trying to hang a mirror. There is only one way to hang a painting, using a nail and a hammer. But there are two ways of hanging a mirror, using a nail and a hammer or using a screw and a driver, but  $j$  is only aware of the former. Agent  $i$  possesses a screw, a screw driver and a hammer, but needs a nail in addition to the hammer to hang the painting. On the other hand,  $j$  possesses a nail, and believes that to hang the mirror, it needs a hammer in addition to the nail. Now, consider the dialogue depicted in Figure 5 (described here in natural language) between the two agents.

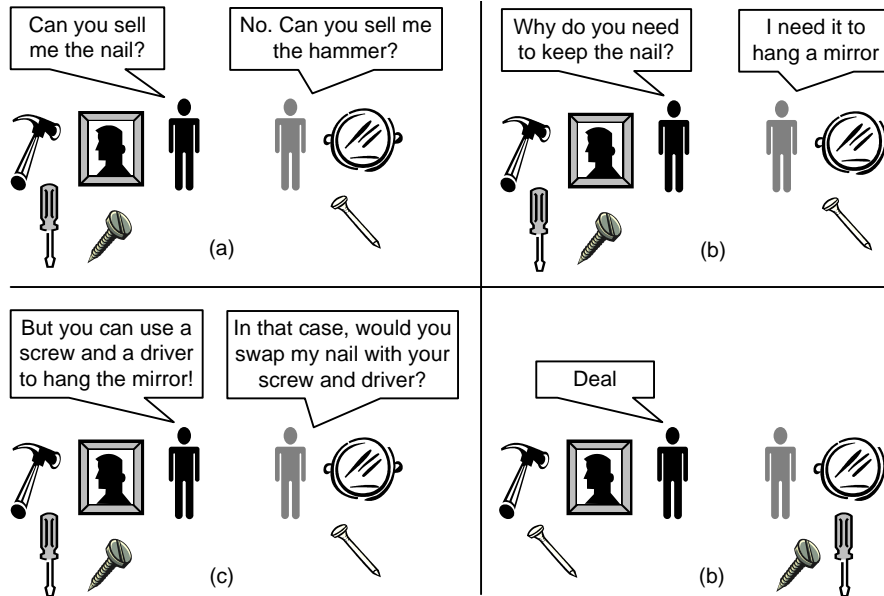


Figure 5: Dialogue between agent  $i$  (black) and  $j$  (gray)

At first,  $j$  was not willing to give away the nail because it needed it to achieve its goal. But after finding out the reason for rejection,  $i$  managed to persuade  $j$  to give away the nail by providing an alternative plan for achieving the latter's goal.

To enable the above dialogue, we need a variant of protocol IBNP1 that enables agents to exchange information about new plans. This is described in IBN Protocol 2 shown in Table 3 (the only difference is the new step 2.d.iii.C).

We formalise the painting/mirror domain in the example below.

**Example 8.** Let the agents' main goal, resources, and goal achievement knowledge be as follows:

- Agent  $i$ :
  - $G(i) = \text{hangpainting}$
  - $\text{sub}(\text{hangpainting}, \{\text{nail}, \text{hammer}\}); \text{sub}(\text{hangmirror}, \{\text{screw}, \text{driver}\})$
  - $\Lambda^0(i) = \{\text{screw}, \text{driver}, \text{hammer}\}$
- Agent  $j$ :
  - $G(j) = \text{hangmirror}$
  - $\text{sub}(\text{hangmirror}, \{\text{nail}, \text{hammer}\})$
  - $\Lambda^0(j) = \{\text{nail}\}$

Given their current knowledge, both agents want the nail and the hammer. Following is a formalisation, in our framework, of the (key elements of) dialogue presented by Parsons et al [25].

1.  $\text{propose}(i, (\Lambda^0, \Lambda^1, p^1))$ , where  $\Lambda^1(i) = \{\text{screw}, \text{driver}, \text{hammer}, \text{nail}\}$ ,  $\Lambda^1(j) = \{\}$ , with some appropriate payment  $p^1$ .
2.  $\text{propose}(j, (\Lambda^0, \Lambda^2, p^2))$ , where  $\Lambda^2(i) = \{\text{screw}, \text{driver}\}$ ,  $\Lambda^2(j) = \{\text{nail}, \text{hammer}\}$ , with some appropriate payment  $p^2$ .
3.  $\text{why}(i, \text{nail})$
4.  $\text{assert}(j, I_j(\text{hangmirror}))$
5.  $\text{achieves}(i, \text{sub}(\text{hangmirror}, \{\text{screw}, \text{driver}\}))$
6.  $\text{propose}(j, (\Lambda^0, \Lambda^3, p^3))$ , where  $\Lambda^3(i) = \{\text{hammer}, \text{nail}\}$ , and  $\Lambda^3(j) = \{\text{screw}, \text{driver}\}$ , with some appropriate payment  $p^3$ .
7.  $\text{accept}(i, (\Lambda^0, \Lambda^3, p^3))$

Note that in the above example, if agents use IBNP1 or a standard bargaining protocol, they could not reach a deal whereby they both believe they can achieve their goals. When using IBNP2 and assuming that their resource valuations allow for an appropriate payment to be found, the agents will indeed be able to reach a deal. This example is showing how IBNP2 allows for qualitative improvement in the outcome of automated negotiation. When qualitative improvement is not possible - for example because IBNP1 or a simple bargaining allow a deal to be found - IBNP2 may still allow for quantitative improvement (i.e. increased agents' utility).

The next proposition formalizes these ideas and states that revealing information about new alternatives can only be beneficial to agents. Let  $\text{sub}()$  be the goal decomposition relation in the system, let  $\text{sub}_i^t()$  be what is known from that relation by agent  $i$  at a given point in time  $t$ .

**Proposition 6.** *At any given time  $t$ , if the step 2.d.III.C of protocol IBNP2 has been played by agent  $j$ , as new tuples are added to the decomposition relation of agent  $i$   $\text{sub}_i^t()$ , the utility of the current offer  $\delta^t$  for agent  $i$  may only increase.*

*Proof.* As new alternatives are added to  $sub_i$ , the set of available alternative plans  $i$  can choose from ( $\mathcal{T}_i^* = \mathcal{T}_{\Lambda(i), dgoals_j^t} \cap \mathcal{T}(G(i))$ ) can only increase. Recall that  $u_i(\Lambda^t(i)) = \max_{T \in \mathcal{T}_i^*} \tilde{u}_i(T)$ . Hence, as  $u_i(\Lambda^t(i))$  is applied to maximise over a monotonically increasing set, its value can increase but not decrease. Consequently,  $u_i(\delta^t)$  is non-decreasing.  $\square$

## 8 Discussion

It is important to note that the analysis presented in this paper is all in the context of *cooperative* negotiation. By this, we mean that while agents seek to accept offers that maximise their individual utilities, they are not *strategic* in the game-theoretic sense. More precisely, we assumed that agents do not lie by misreporting their goals or their, or providing false advice about the possible ways to achieve a goal. This amounts to assuming that there is an external mechanism that can verify whether such statements are true (e.g. by observing the agent’s actions to monitor their plan execution) and providing a penalty high enough for lying to be ruled out. For example, a centralised taxi dispatch system might require decentralised negotiation between individual taxis (e.g. to exchange jobs based on location suitability), but be able to monitor their subsequent execution against the agreement reached. Of course, this assumption does not hold in all domains, and particularly in completely decentralised domains in which agents have no mechanism for monitoring each others’ plan execution. Having said that, an extensive game theoretic analysis is beyond the scope of the present paper, and will be an interesting venue for future work.

Another issue worth discussing is what happens when the conditions that guarantee our results are not met. As expected, in such situations, IBN does *not* necessarily always offer an advantage over bargaining protocols when it comes to reaching a deal. But this may not be such a bad thing, as we will explain below.

In this context, we are mainly interested in cases with *negative* interaction between goals. Consider the case in which one of agent  $i$ ’s goals  $g$  invalidates one of agent  $j$ ’s goals  $g'$ . Here, by revealing its goal, agent  $i$  risks causing its offer’s utility to *decrease* from the point of view of agent  $j$ . In particular, this may lead  $j$ ’s complete plan to be invalidated (turning it into a partial plan) since  $g'$  is no longer achievable.

Having said that, although the utility of a particular offer decreases, this is not necessarily a bad thing. Indeed, discovering negative interaction during the negotiation phase is better than discovering it during plan execution. In other words, IBN acts as a mechanism for early detection of execution-time conflicts, as in the hierarchical plan merging literature [8]. Notably, in the context of negotiation, this has implications on the agents’ motivation to share information. That is, even if agents do not *lie* by claiming to have goals they do not actually have, agents may still be able to *deceive* by deliberately withholding information goals that interact negatively with their opponent’s goals. This may give the (false) impression that a particular offer is acceptable. We are interested in exploring the strategic implications of this in future work.

## 9 Related Work

Below, we discuss some related work. We first discuss how our work relates to other ABN frameworks. Then, we discuss the connection between our work and work on negotiation in state-oriented domains [36]. Finally, we draw connections to the hierarchical plan merging literature.

### 9.1 Other ABN Frameworks

A variety of logic-based argument-based negotiation frameworks have been presented in the literature. They focus on different knowledge representation issues, and only present specific examples that show how negotiation can lead to agreement. For example, Parsons et al [25] present a framework based on the logic-based argumentation framework of Elvang-Gøransson et al [11]. The framework of Sadri et al [37] uses abductive logic programming [15]. Rahwan [28] uses a variant of Amgoud and Cayrol’s preference-based argumentation framework [2]. Recently, Amgoud et al introduced a general model of argument-based negotiation [3], although it focuses on the connection between arguments and deals in general, and does not have a specific notion of goal revelation.

However, in all of the above frameworks, there is no high-level analysis of the outcomes of goal revelation. This paper presents the first precise analysis of outcomes of goal revelation in these kinds of negotiation frameworks. To our knowledge, the only such analysis is our own preliminary exploration [30], which did not capture all aspects of IBN (e.g. goal hierarchies) and did not analyse any specific IBN protocol.

It is important to note that the framework presented in this paper is not an *alternative* to existing ABN frameworks. Instead, it provides means to analyse a particular aspect of ABN frameworks, that of goal revelation and its effect on negotiation outcomes. In other words, the contribution of this paper are aimed at complementing our understanding of those ABN frameworks in which goal revelation takes place, and the conditions under which such goal revelation is guaranteed to be beneficial.

Indeed, the ABN literature contains frameworks that allow the exchange of a variety of other kinds of information, such as threats and rewards [21, 4, 35], tips and warnings [1], and so on. Such other kinds of arguments are beyond the scope of the present work, although opportunities do exist to combine them with our work in the future.

### 9.2 State-Oriented Domains

There are some similarities between our analysis and Rosenschein and Zlotkin’s analysis of State-Oriented Domains (SOD) [36]. A SOD is defined in terms of a set of states  $\mathcal{S}$ , and agent  $i$ ’s goal is defined as a set  $G_i \subseteq \mathcal{S}$  of desirable states with a fixed worth  $w_i$ . In the case of two agents  $i$  and  $j$ , a deal  $\delta$  is a ‘joint plan’ which identifies a set of actions for each agent to perform, and maps the current state  $s$  to a new state in  $G_i \cap G_j$  in the non-conflict case. Hence, the utility derived by agent  $i$  from deal  $\delta$  is the difference between the worth of its goal and the cost of  $i$ ’s role in the agreed joint plan, formally (in their language):  $\text{Utility}_i(\delta) = w_i - c_i(\delta)$ .

While this formulation is very similar to ours (see Definitions 13 and 14), a fundamental difference between SODs and our model is that deals in SODs are fully specified joint plans. This contrasts with our model, in which deals are re-allocations of resources, regardless of how these resources are used by individual agents (i.e. which plans end up being executed by the agents). This means that in SODs, agents can accurately identify the final state, and thus each agent can fully determine (from information in a given offer) the positive side-effects of the other agents' actions on its own goal. In our model, on the other hand, agents are involved in resource (re-)allocation [12, 5], and are unable to accurately identify towards which goals these resources will be used by others. IBN enables an exploration of these underlying goals in the context of resource exchange. In other words, the advantage of IBN lies in its ability to deal with the uncertainty about how the exchanged resources are being used towards underlying goals (IBNP1) as well as its ability to deal with the incompleteness of the information concerning alternative ways to achieve these goals (IBNP2).

Note that some game-theoretic approaches deal with incomplete information via so-called direct-revelation mechanisms [24]. Agents reveal all private information at the beginning, and this information is then used by a centralised mechanism to determine the outcome/deal. The mechanism is often designed in such a way that revealing private information truthfully (i.e. the agent's true type) is the optimal strategy. For example, Rosenschein and Zlotkin [36] present a direct revelation mechanism for SODs with mixed (probabilistic) deals. However, our domain is different in that negotiation is over resources rather than actions, and in that the values of resources are private and subjective. To our knowledge, no direct revelation mechanism has been identified for our setting, and it is indeed unclear if one exists. Furthermore, the direct revelation mechanisms are associated with a number of drawbacks, as discussed below.

First, the revelation of all the information can be computationally very expensive [7]. Indeed, the revelation phase consists of all agents revealing all the unshared information. This clearly dissolves the inherent benefit and realism of distributed systems (such as multi-agent systems) in which the information is and has to stay distributed. IBN allows agents to progressively reveal their goals, and potentially reach agreement without full revelation.

Secondly, in principled negotiation approaches [19, 6], it is important that agents minimise the amount of information they reveal about their preferences since such revelation may weaken their positions (e.g. incentive not to reveal their reservation price) [27, 34].

Furthermore, it has been shown that humans tend to minimise the amount of private information they reveal during negotiation [18]. Such considerations are important in an increasing number of human-agent negotiation models [17, 16, 23]. This minimality also serves the purpose of reducing communication complexity induced by direct revelation.

Another advantage of our framework when compared with SODs is that it allows for side payments, which enables a wider range of possible deals [12]. For example, payment enables buyer-seller scenarios which are common in electronic commerce.

### 9.3 Hierarchical Plan Merging

Finally, it is worth noting that our work differs from multi-agent hierarchical plan merging [8], which assumes agents are fully aware of each other’s goals, and uses a centralised mechanism for identifying synergies between plans. We depart from a position where agents have no knowledge of each other’s goals. And while the objective of hierarchical coordination research is on finding optimal ways to maximise positive interaction among the goals of *cooperative* agents, our aim is to explore interaction among self-interested agents who may not be willing to share information about their goals, unless sharing such information benefits them.

## 10 Conclusion

This paper provides new insights on existing research on logic-based argumentation-based negotiation. While much has been said about the intuitive advantage of argument-based negotiation over other forms of negotiation, very little has been done on making these intuitions precise. We started bridging this gap by characterising exactly how the set of reachable deals expands as agents progressively explore each other’s underlying goals and alternative ways to achieve them. We presented two specific protocols and formally showed conditions under which they are guaranteed to improve the outcome of the negotiation, namely when goals can interact positively.

There is a significant gap between argumentation-based models of negotiation and those analysed in the game-theoretic literature. Bridging this gap completely is beyond the scope of any single paper (indeed, connections between argumentation and game theory are still at a very early stage [29]). Having said that, the formal framework presented in the paper begins to make the connection between arguments on one hand, and outcomes of negotiation on the other hand, more explicit. We hope that this will enable subsequent game-theoretic analysis of ABN models, which is crucial for addressing issues such as deception and lying.

The formal framework presented in the paper provides a foundation for elaborate analysis of a variety of IBN protocols, other than those addressed in this paper. For example, it would be interesting to analyse symmetric protocols that allow simultaneous goal revelation.

Another direction of future research is exploring the case of negative interaction (i.e. interference) among agents’ goals. In this paper, conflict between goals was implicit in the sense that multiple goals may compete over the same resources, and thus conflict was explicit only at the resource level. However, in some domains, a goal may hinder the achievement of another goal (e.g. as in the TAEMS coordination framework [22]). In such cases, exploring underlying interests can reveal hidden conflicts that would otherwise only be discovered at run-time (i.e. when agents execute their plans) and prevent agents from achieving their goals despite reaching an acceptable deal. Moreover, agents may not wish to disclose their goals, since these may deter their counterparts from agreeing on certain deals. Thus, one would have to explore the trade-off between the potential benefit and potential loss in revealing goals. In a related context, the strategic issues associated with non-sincere agents and the possibility of

agents lying about their goals opens up many game-theoretic questions.

Another direction of future research is investigating our framework empirically. In other work [26] we began an investigation of a variant of this framework comparing a variety of bargaining and interest-based strategies.

It is worthwhile mentioning that there is extensive literature on deal reachability in alternating-offer protocols in which individual rational offers are traversed (see for example [12, 10, 9]). It will be interesting to explore the reachability of deals with IBN protocols under different conditions in a similar manner.

## Acknowledgement

This work is partially supported by the Australian Research Council, Discovery Grant DP0557487. The authors acknowledge the anonymous AAAI reviewers for their useful comments.

## References

- [1] L. Amgoud, J.-F. Bonnefon, and H. Prade. The logical handling of threats, rewards, tips, and warnings. In K. Mellouli, editor, *Symbolic and Quantitative Approaches to Reasoning with Uncertainty, 9th European Conference, ECSQARU 2007, Hammamet, Tunisia, October 31 - November 2, 2007, Proceedings*, pages 235–246, 2007.
- [2] L. Amgoud and C. Cayrol. A reasoning model based on the production of acceptable arguments. *Annals of Mathematics and Artificial Intelligence*, 34(1–3):197–215, 2002.
- [3] L. Amgoud, Y. Dimopoulos, and P. Moraitis. A unified and general framework for argumentation-based negotiation. In *AAMAS '07: Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems*, New York, NY, USA, 2007. ACM.
- [4] L. Amgoud and H. Prade. Handling threats, rewards and explanatory arguments in a unified setting. *International Journal Of Intelligent Systems*, 20(12):1195–1218, 2005.
- [5] Y. Chevaleyre, P. E. Dunne, U. Endriss, J. Lang, M. Lemaître, N. Maudet, J. Padget, S. Phelps, J. A. Rodríguez-Aguilar, and P. Sousa. Issues in multiagent resource allocation. *Informatica*, 30:3–31, 2006.
- [6] S. Cohen. *Negotiating Skills for Managers*. MacGraw-Hill, New York, 2002.
- [7] V. Conitzer and T. Sandholm. Computational criticisms of the revelation principle. In *Proceedings of the 5th ACM Conference on Electronic Commerce (EC-04)*, pages 262–263, 2004.

- [8] J. S. Cox and E. Durfee. Discovering and exploiting synergy between hierarchical planning agents. In J. Rosenschein, T. Sandholm, M. J. Wooldridge, and M. Yokoo, editors, *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS-2003)*, pages 281–288. ACM Press, 2003.
- [9] P. E. Dunne and Y. Chevaleyre. The complexity of deciding reachability properties of distributed negotiation schemes. *Theoretical Computer Science*, 396(1-3):113–144, 2008.
- [10] P. E. Dunne, M. Wooldridge, and M. Laurence. The complexity of contract negotiation. *Artificial Intelligence*, 164(1-2):23–46, 2005.
- [11] M. Elvang-Gøransson, P. Krause, and J. Fox. Dialectic reasoning with inconsistent information. In D. Heckerman and A. Mamdani, editors, *Proceedings of the 9th Conference on Uncertainty in Artificial Intelligence*, pages 114–121, Washington D. C., USA, 1993. Morgan Kaufmann.
- [12] U. Endris, N. Maudet, F. Sadri, and F. Toni. Negotiating socially optimal allocations of resources. *Journal of artificial intelligence research*, 25:315–348, 2006.
- [13] K. Erol, J. Hendler, and D. Nau. Semantics for hierarchical task network planning. Technical Report CS-TR-3239, UMIACS-TR-94-31, Department of Computer Science, University of Maryland, 1994.
- [14] P. Faratin, C. Sierra, and N. R. Jennings. Using similarity criteria to make trade-offs in automated negotiations. *Artificial Intelligence*, 142(2):205–237, 2002.
- [15] T. Fung and R. Kowalski. The IFF proof procedure for abductive logic programming. *Journal of Logic Programming*, 33(1):151–165, 1997.
- [16] Y. Gal and A. Pfeffer. Modeling reciprocity in human bilateral negotiation. In *National Conference on Artificial Intelligence (AAAI)*, Vancouver, British Columbia, 2007.
- [17] B. Grosz, S. Kraus, S. Talman, B. Stossel, and M. Havlin. The influence of social dependencies on decision-making: Initial investigations with a new game. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems - Volume 2 (AAMAS 2004)*, pages 294–301, New York, NY, USA, 2004. IEEE Computer Society.
- [18] P. Heiskanen, H. Ehtamo, and R. Hamalain. Constraint proposal method for computing Pareto solutions in multi-party negotiations. *European Journal of Operational Research*, 133(1):44–61, 2001.
- [19] J. Hiltrop and S. Udall. *The Essence of Negotiation*. Prentice Hall, Englewood Cliffs, NJ, 1995.
- [20] N. R. Jennings, P. Faratin, A. R. Lomuscio, S. Parsons, C. Sierra, and M. J. Wooldridge. Automated negotiation: prospects, methods and challenges. *International Journal of Group Decision and Negotiation*, 10(2):199–215, 2001.

- [21] S. Kraus, K. Sycara, and A. Evenchik. Reaching agreements through argumentation: A logical model and implementation. *Artificial Intelligence*, 104(1–2):1–69, 1998.
- [22] V. Lesser, K. Decker, T. Wagner, N. Carver, A. Garvey, B. Horling, D. Neiman, R. Podorozhny, M. NagendraPrasad, A. Raja, R. Vincent, P. Xuan, and X. Zhang. Evolution of the GPGP/TAEMS Domain-Independent Coordination Framework. *Autonomous Agents and Multi-Agent Systems*, 9(1):87–143, July 2004.
- [23] R. Lin, S. Kraus, J. Wilkenfeld, and J. Barry. An automated agent for bilateral negotiation with bounded rational agents with incomplete information. In *Proc. of the 17th European Conference on Artificial Intelligence (ECAI)*, pages 270–274, 2006.
- [24] A. Mas-Colell, M. D. Whinston, and J. R. Green. *Microeconomic Theory*. Oxford University Press, New York NY, USA, 1995.
- [25] S. Parsons, C. Sierra, and N. Jennings. Agents that reason and negotiate by arguing. *Journal of Logic and Computation*, 8(3):261–292, 1998.
- [26] P. Pasquier, R. Hollands, F. Dignum, I. Rahwan, and L. Sonenberg. An empirical study of interest-based negotiation. In M. L. Gini, R. J. Kauffman, D. Sarppo, C. Dellarocas, and F. Dignum, editors, *Proceedings of the Ninth International Conference on Electronic Commerce (ICEC), University of Minnesota, Minneapolis, USA*, pages 339–348, New York NY, USA, 2007. ACM Press.
- [27] D. G. Pruitt. *Negotiation Behavior*. Academic Press, New York, USA, 1981.
- [28] I. Rahwan. *Interest-based Negotiation in Multi-Agent Systems*. PhD thesis, Department of Information Systems, University of Melbourne, Melbourne, Australia, 2004.
- [29] I. Rahwan and K. Larson. Mechanism design for abstract argumentation. In L. Padgham, D. Parkes, J. Mueller, and S. Parsons, editors, *7th International Joint Conference on Autonomous Agents & Multi Agent Systems, AAMAS’2008, Estoril, Portugal*, pages 1031–1038, 2008.
- [30] I. Rahwan, P. McBurney, and L. Sonenberg. Bargaining and argument-based negotiation: Some preliminary comparisons. In I. Rahwan, P. Moraitis, and C. Reed, editors, *Argumentation in Multi-Agent Systems, First International Workshop, ArgMAS 2004, New York, NY, USA, July 19, 2004, Revised Selected and Invited Papers*, volume 3366 of *Lecture Notes in Artificial Intelligence*, pages 176–191, Berlin, Germany, 2004. Springer Verlag.
- [31] I. Rahwan, P. Pasquier, L. Sonenberg, and F. Dignum. On the benefits of exploiting underlying goals in argument-based negotiation. In R. C. Holte and A. Howe, editors, *Proceedings of the 22nd AAAI Conference on Artificial Intelligence (AAAI-2007)*, pages 116–121, Menlo Park CA, USA, 2007.

- [32] I. Rahwan, S. D. Ramchurn, N. R. Jennings, P. McBurney, S. Parsons, and L. Sonenberg. Argumentation based negotiation. *Knowledge Engineering Review*, 18(4):343–375, 2003.
- [33] I. Rahwan, L. Sonenberg, and F. Dignum. Towards interest-based negotiation. In J. Rosenschein, T. Sandholm, M. J. Wooldridge, and M. Yokoo, editors, *Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2003)*, pages 773–780. ACM Press, 2003.
- [34] H. Raiffa. *The Art and Science of Negotiation*. Harvard University Press, Cambridge MA, USA, 1982.
- [35] S. D. Ramchurn, C. Sierra, L. Godo, and N. R. Jennings. Negotiating using rewards. *Artificial Intelligence*, 171(10–15):805–837, 2007.
- [36] J. Rosenschein and G. Zlotkin. *Rules of Encounter: Designing Conventions for Automated Negotiation among Computers*. MIT Press, Cambridge MA, USA, 1994.
- [37] F. Sadri, F. Toni, and P. Torroni. Logic agents, dialogues and negotiation: an abductive approach. In K. Stathis and M. Schroeder, editors, *Proceedings of the AISB 2001 Symposium on Information Agents for E-Commerce*, 2001.
- [38] T. W. Sandholm. Distributed rational decision making. In G. Weiss, editor, *Multiagent Systems: A Modern Approach to Distributed Artificial Intelligence*, pages 201–258. The MIT Press, Cambridge, MA, USA, 1999.
- [39] E. Wolfstetter. Auctions: An introduction. *Journal of Economic Surveys*, 10:367–420, 1996.

## Appendix: Semantics of Protocols

In this appendix, we provide a semantics of the protocols IBNP1 and IBNP2 in the form of finite-state-machines. This finite-state-machine approach is very common for the description of simple protocols, and has been used in the context of argumentation by Parsons et al [25]. The semantics of protocol IBNP1 is shown in Figure 6. Here, the process starts (State 0) when one agent  $i$  makes a proposal to another agent  $j$ , denoted by  $\text{propose}(i, \delta^t)$ . This moves the protocol to State 1a, in which the agent  $j$  may counter propose (leading to the mirror State 1b with roles reversed), accept (leading to terminal State 2), reject (leading to terminal State 3), or initiate an IBN sub-protocol by uttering  $\text{why}(j, x)$  for some goal  $x$  declared by agent  $i$  (State 4a). From here, agent  $i$  can either assert a super-goal or decline, leading to State 5a. From here, agent  $j$  either asks for another super-goal, leading back to State 4a, accept, or pass the IBN phase. This finite-state-machine specification provides a complement to the specification shown in Table 2 by showing how the utterances transform the dialogue from one state to another, and specifying exactly when the dialogue terminates.

The semantics of protocol IBNP2 has an almost identical finite state machine, so we do not include it separately here. It suffices to say that an additional transition,

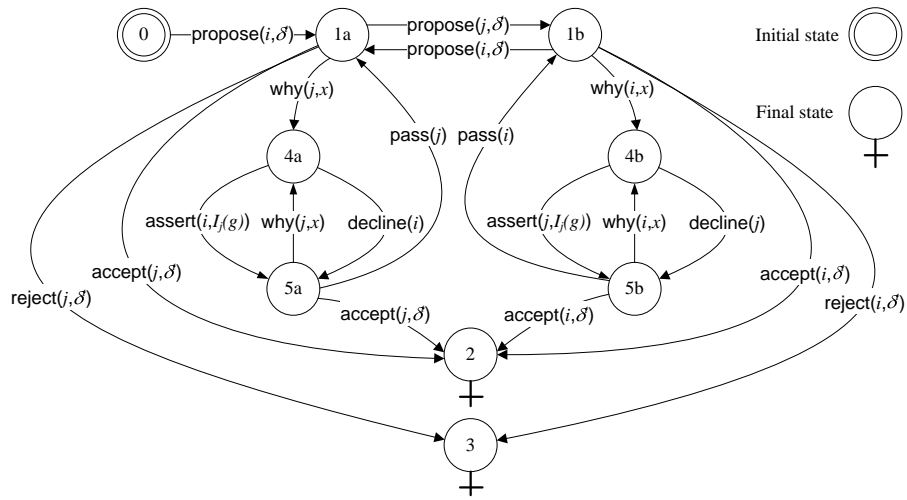


Figure 6: Semantics of protocol IBNP1 as a finite state machine

labelled  $\text{achieves}(j, \text{sub}(g, \{x_1, \dots, x_n\}))$ , going from State 5a to State 1b. A symmetric transition is also added from State 5b to State 1a.