On the Benefits of Exploiting Underlying Goals in Argument-based Negotiation

Iyad Rahwan
Institute of Informatics
British University in Dubai
P.O.Box 502216, Dubai, UAE
(Fellow) School of Informatics
University of Edinburgh, UK

Philippe Pasquier, Liz Sonenberg
Dept. of Information Systems
University of Melbourne
Parkville, VIC 3010 Australia

Frank Dignum
Department of Information
& Computing Sciences
Utrecht University
Utrecht, The Netherlands

Abstract

Interest-based negotiation (IBN) is a form of negotiation in which agents exchange information about their underlying goals, with a view to improving the likelihood and quality of a deal. While this intuition has been stated informally in much previous literature, there is no formal analysis of the types of deals that can be reached through IBN and how they differ from those reachable using (classical) alternating offer bargaining. This paper bridges this gap by providing a formal framework for analysing the outcomes of IBN dialogues, and begins by analysing a specific IBN protocol.

Introduction

Negotiation is a form of interaction in which a group of agents, with conflicting interests, try to come to a mutually acceptable agreement on the division of scarce resources. Approaches to automated negotiation can be classified to those based on (1) auctions; (2) bargaining; and (3) argumentation. A common aspect of auction and bilateral bargaining approaches is that they are *proposal-based*. 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 (Rahwan *et al.* 2003). This paper is concerned with a particular style of argument-based negotiation, namely *interest-based negotiation* (IBN) (Rahwan, Sonenberg, & Dignum 2003), 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 (Parsons, Sierra, & Jennings 1998)). 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.[1] We 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 a simple IBN protocol and 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. As such, the paper bridges the gap between the theory and practice of ABN, and provides a key first step towards understanding the dynamics of more complex IBN dialogues.

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 $\mathcal{A}$ (Endris *et al.* 2006).

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.

[1] This abstraction is common and has been used in the context of automated planning (Erol, Hendler, & Nau 1994) and multi-agent coordination (Cox & Durfee 2003).

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 (Endris *et al.* 2006).

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

Let Δ be the set of all possible deals. By overloading the notion of utility, 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* δ *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* δ *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.[2] We denote by $\Delta^* \subseteq \Delta$ the *set of individual rational deals*.

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 alternative-offer 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. 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 every possible allocation of resources, but also every possible payment. Various computational frameworks for bargaining have been proposed in order to enable agents to reach deals quickly. For example, Faratin et al

[2]This is equivalent to saying that the new allocation *Pareto dominates* the initial allocation, given the payment.

Bargaining Protocol 1 (BP1):
Agents start with resource allocation Λ^0 at time $t = 0$
At each time $t > 0$:

1. propose(i, δ^t): Agent i proposes to j deal $\delta^t = (\Lambda^0, \Lambda^t, p^t)$ which has not been proposed before;
2. Agent j either:
   (a) accept(j, δ^t): accepts, and negotiation terminates with allocation Λ^t and payment p^t; or
   (b) reject(j, δ^t): rejects, and negotiation terminates with allocation Λ^0 and no payment; or
   (c) makes a counter proposal by going to step 1 at the time step $t + 1$ with the roles of agents i and j swapped.

Table 1: Basic bargaining protocol

(Faratin, Sierra, & Jennings 2002) 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).[3]

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*

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

where $t = 1, \ldots, n$ *represents time.*

Definition 9. ***(Protocol-Reachable Deal)*** *Let* P *be a protocol. A deal* δ^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* δ^t *is offered by some agent at time* t *and* δ^t *is individual rational given* u_i^t, u_j^t*.*

Underlying Interests

In most existing alternating-offer bargaining negotiation frameworks, agents' utility functions are assumed to be *predetermined* (e.g. as weighted sums) and *fixed* throughout the interaction. That is, throughout the dialogue history, $u_i^1 = \cdots = u_i^n$ for any agent i.

We now present a framework for capturing the interdependencies between goals at different levels of abstraction.[4]

Let $\mathcal{G} = \{g_1, \ldots, g_m\}$ be the 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, $sub(g, \{g_1, \ldots, g_n\})$ means that achieving all the goals $g_1, \ldots, g_n$ results in achieving the higher-level goal g. Each sub-goal in the set $\{g_1, \ldots, g_n\}$ 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

[3]To enable studying changes in the utility function later in the paper, we will superscript utility functions with time-stamps.

[4]Although this framework is simpler than those in the planning literature, its level of abstraction is sufficient for our purpose.

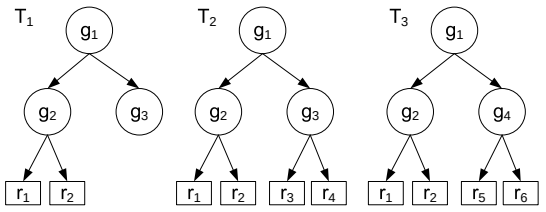

Figure 1: Partial plans (T_1) and complete plans (T_2, T_3)

tree (called *goal tree* or *plan*). This condition is reasonable since the sub-goal relation captures specialisation of abstract goals into more concrete goals.

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 $x_1, \ldots, x_n \in \mathcal{G} \cup \mathcal{R}$ such that $sub(g, \{x_1, \ldots, x_n\})$;*[5]
- *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 goals $\mathcal{G} = \{g_1, \ldots, g_4\}$ and $\mathcal{R} = \{r_1, \ldots, r_6\}$ 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.

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

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.

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 $T \in \mathcal{T}$ 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*.[6] We further assume that agent i assigns a *worth* to this goal $worth_i(G(i)) \in \mathbb{R}$.

[5] I.e. among alternatives for achieving g, only one is selected.

[6] Multiple goals can be expressed by a single goal that has one possible decomposition.

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 it 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 $\breve{T}$. 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:*

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

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 $v_i : \mathcal{R} \to \mathbb{R}$ 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: $cost_i(T) = \sum_{r \in rleaves(T)} v_i(r)$. Then, we can define the *utility of plan* as follows.[7]

Definition 14. ***(Utility of a Plan)*** *Let i be an agent with goal $G(i)$ and resources $\Lambda(i)$. 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 $\tilde{u}_i : \mathcal{T}_i^* \to \mathbb{R}$ is defined as follows:*

$$\tilde{u}_i(T) = \begin{cases} 0 & \text{if } T = \breve{T}, \\ worth_i(G(i)) - cost_i(T) & \text{otherwise} \end{cases}$$

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 simply the sum of the values of those resources.

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 $worth_i(g_1) = 85$ 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:*
$\tilde{u}_i(T_2) = 85 - (20 + 10 + 6 + 5) = 44$
$\tilde{u}_i(T_3) = 85 - (20 + 10 + 8 + 7) = 40$
$\tilde{u}_i(\breve{T}) = 0$

[7] 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.

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.

Definition 15. ***(Utility)*** *The* utility *of agent $i \in \mathcal{A}$ is defined as a function $u_i : 2^{\mathcal{R}} \rightarrow \mathbb{R}$ such that:*

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

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.*

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*[8] 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.

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.

For the time being, we assume no negative interaction among goals.[9] In other words, the achievement of one goal does not hinder the achievement of another.

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 $cgoals_i^t$ at time t is:*

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

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 plan $I_j(T_4)$. 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.*

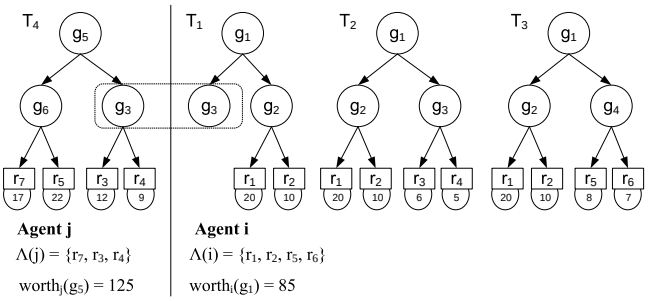

Figure 2: Agent i can benefit from j's committed goal

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

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 $\Lambda(i)$, 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 $\tilde{u}_i(\breve{T}) = 0$ (recall that $T_2 \notin \mathcal{T}_i^*$ for now). Therefore, $u_i(\Lambda(i)) = 70$. This contrasts with the calculation that does not take j's goal into account, in which case $u_i(\Lambda(i)) = 40$.*

Case Study: An IBN Protocol

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 underling goals symmetrically (e.g. simultaneously) or asymmetrically, etc. We now look at a specific IBN protocol and analyse it using the above concepts.

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

[8] 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.

[9] In this paper, *negative* interaction among goals is only captured through the overlap of resources needed by two goals. We do not address explicit interference among goals.

that this protocol is asymmetric, since during the IBN subdialogue, 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.

IBN Protocol 1 (IBNP1):
Agents start with resource allocation Λ^0 at time $t = 0$

At each time $t > 0$

1. propose(i, δ^t): Agent i proposes to j deal $\delta^t = (\Lambda^0, \Lambda^t, p^t)$ which has not been proposed before;
2. Agent j either:
   (a) accept(j, δ^t): accepts, and negotiation terminates with allocation Λ^t and payment p^t; or
   (b) reject(j, δ^t): rejects, and negotiation terminates with allocation Λ^0 and no payment; or
   (c) makes a counter proposal by going to step 1 at the next time step with the roles of agents i and j swapped; or
   (d) switches to interest-based dialogue on δ^t. Let $dgoals_i^t = \emptyset$ for all $i \in \mathcal{A}$ be each agents' declared goals.
      i. why(j, x): j asks i for underlying goal for a resource or declared goal $x \in \Lambda^t(i) \cup dgoals_i^t$;
      ii. i either:
         A. assert$(i, I_i(g))$: i responds by stating a goal, which is added to $dgoals(i)$; or
         B. decline(i): declines giving the information;
      iii. j either:
         A. accept(j, g): j accepts δ^t, if now more favourable; or
         B. seeks more information by going to step 2.d.i; or
         C. pass(j): j skips its turn, moving the protocol to step 2 with i taking the role of deciding what to do next.

Table 2: A simple IBN protocol

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 $\breve{T}$. 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:*

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 some resource, say r_3, and the following follows:

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

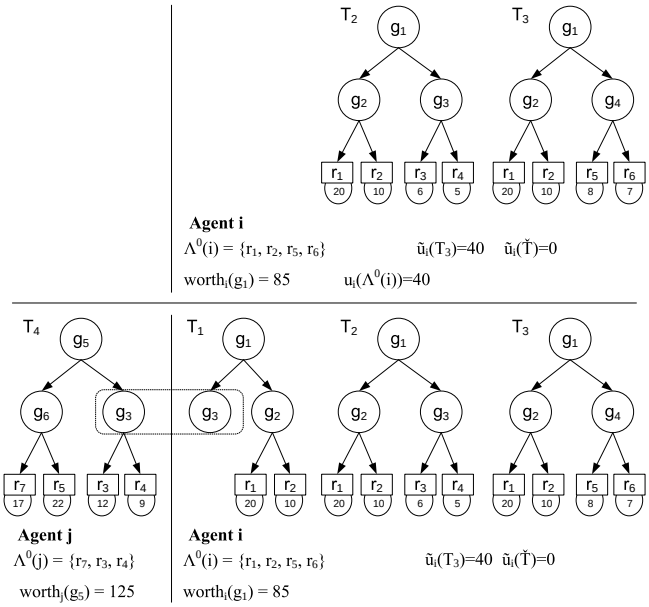

Figure 3: Different stages of an IBN dialogue

At this point, i would be willing to give up r_3 and r_4, since plan T_1 now becomes a viable option for i. 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. 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 δ 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.

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. □

We are mainly interested in how agents' perceptions of the utility of allocations changes 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.

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

The below proposition then follows:

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 $dgoals_i^t \subseteq cgoals_i^t$. □

This 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 $\mathcal{T}_i^* = \mathcal{T}_{\Lambda(i),dgoals_j^t} \cap \mathcal{T}(G(i))$.

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 δ^t (Step 1). By definition, $leaves(T_i^t) \subseteq \Lambda^t(i)$, i.e. i must obtain through δ^t every resource needed for achieving T_i^t. After this request (Step 2.d), j could ask why(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, $dgoals_i^t$ will contain the set of goals that are immediate parents of resources $r \in leaves(T_i^t)$. 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 $dgoals_i^t$. □

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. □

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 $\Lambda^t(i)$.

Conclusion

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 began bridging this gap by characterising exactly how the set of reachable deals expands as agents progressively explore each other's underlying goals. We also presented one specific protocol and showed how it provides one useful way to exchange information about goals.

This paper opens many future possibilities. Although the protocol analysed here is simple, the paper presents a step towards more elaborate analysis of a variety of other IBN protocols (e.g. symmetric ones). Another direction of future research is exploring the case of negative interaction (i.e. interference) among agents' goals. In such cases, agents may not wish to disclose their goals, since this could reduce the likelihood or quality of deals. One would have to explore the trade-off between the potential benefit and potential loss in revealing goals. Finally, the possibility of agents lying about their goals opens up many game-theoretic questions.

It is worth noting that our work differs from multi-agent hierarchical plan merging (Cox & Durfee 2003), which assume agents are fully aware of each other's goals. 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.

Acknowledgement

This work is partially supported by the Australian Research Council, Discovery Grant DP0557487.

References

Cox, J. S., and Durfee, E. 2003. Discovering and exploiting synergy between hierarchical planning agents. In *Proc. AAMAS*, 281–288. ACM Press.

Endris, U.; Maudet, N.; Sadri, F.; and Toni, F. 2006. Negotiating socially optimal allocations of resources. *Journal of artificial intelligence research* 25:315–348.

Erol, K.; Hendler, J.; and Nau, D. 1994. Semantics for hierarchical task network planning. Technical Report CS-TR-3239, UMIACS-TR-94-31, Department of Computer Science, University of Maryland.

Faratin, P.; Sierra, C.; and Jennings, N. R. 2002. Using similarity criteria to make trade-offs in automated negotiations. *Artificial Intelligence* 142(2):205–237.

Parsons, S.; Sierra, C.; and Jennings, N. 1998. Agents that reason and negotiate by arguing. *Journal of Logic and Computation* 8(3):261–292.

Rahwan, I.; Ramchurn, S. D.; Jennings, N. R.; McBurney, P.; Parsons, S.; and Sonenberg, L. 2003. Argumentation based negotiation. *Knowledge Engineering Review* 18(4):343–375.

Rahwan, I.; Sonenberg, L.; and Dignum, F. 2003. Towards interest-based negotiation. In *Proc. AAMAS*, 773–780. ACM Press.