Reprinted from IJCAI-89, pages 912–917

Negotiation and Task Sharing Among Autonomous Agents in Cooperative Domains

Gilad Zlotkin
Jeffrey S. Rosenschein
Computer Science Department
Hebrew University
Givat Ram, Jerusalem, Israel

Abstract

Research in Distributed Artificial Intelligence is concerned with how automated agents can be designed to interact effectively. One important capability that could aid inter-agent cooperation would be that of negotiation: agents could be built that are able to communicate their respective desires and compromise to reach mutually beneficial agreements.

This work uses the language of game theory to analyze negotiation among automated agents in cooperative domains. However, while game theory generally deals with negotiation in continuous domains and among agents with full information, this research considers discrete domains and the case where agents have only partial information, assumptions of greater interest for artificial intelligence.

A novel, stable, negotiation protocol is introduced for the case of agents who are able to share a discrete set of tasks with one another. The case of agents who may lie to one another during the negotiation, either by hiding some of their tasks or by creating fictitious tasks, is analyzed; it is shown that under some conditions lies are beneficial and "safe," i.e., undiscoverable, while under other circumstances, lies can never be safe.

1 Introduction

1.1 The Negotiation Problem

Research in distributed artificial intelligence (DAI) is concerned with how automated agents can be designed to interact effectively. One important capability that could aid inter-agent cooperation would be that of negotiation: agents could be built that are able to communicate their respective desires and compromise to reach mutually beneficial agreements. While the general concept of "negotiation" has repeatedly been discussed in the artificial intelligence community, there has not been a common vocabulary for analyzing what is meant by the term, nor a developed theory for how automated agents might be made into capable negotiators.

1.2 Previous Work in Distributed AI

Work in DAI has, since its earliest years, been concerned with negotiation strategies. Smith's work on the Contract Net [Smith, 1978] introduced a form of simple negotiation among cooperating agents, with one agent announcing the availability of tasks and awarding them to other bidding agents. Malone refined this technique considerably by overlaying it with a more sophisticated economic model [Malone *et al.*, 1988], proving optimality under certain conditions. While Smith's original work assumed some autonomy among agents, these agents willingly bid for tasks without explicit motivation. Malone's research introduced a motivational framework in the language of economic theory, and at the same time provided a more secure theoretical language in which to discuss the task-sharing algorithm.

Other research in DAI relating to negotiated agreements includes that of Sycara [Sycara, 1988], who modeled labor negotiations from a cognitive standpoint, and Durfee [Durfee, 1988], who introduced negotiation as a key issue in the successful interaction of network nodes in the vehicle monitoring domain.

1.3 Relation to Game Theory

This paper imports game theoretic techniques into an analysis of multi-agent negotiation, in a way analogous to Malone's introduction of economic theory to the Contract Net. By introducing the formal language of game theory, suitably modified for use in an AI context, we can provide tools for designing negotiation into automated agents.

This research follows in the footsteps of [Rosenschein and Genesereth, 1985], which also used certain game-theoretic techniques to model negotiation. There, however, the process of negotiation was severely restricted (the agents could only make single, simultaneous offers); the primary point of that work was to show how varying the axioms of rationality led to altered behavior among agents.

Here, we are not altering the definitions of rationality—we make use of standard game theory definitions for the most part. Instead, we are using game theory insights to analyze problems that are of specific interest to AI, and jettisoning game theory assumptions that are not relevant to AI. For example, game theory negotiation deals with continuous domains; we analyze a discrete negotiation domain, which is more relevant for the case of automated agents negotiating over sets of actions that are to be shared. Similarly, while game theory traditionally deals with agents who have full information, we analyze the case where agents may have partial information, and thus might consider lying to one another during the negotiation process.

2 Definitions and Assumptions

2.1 The Postmen Problem

Two agents A and B have to deliver letters to mailboxes. Each has a set of addressed letters, and wants each letter in his set to be in the mailbox to which it is addressed; after the letters are delivered, the agent must be back in the post office. The only operation with any cost is walking from one place to another (a one meter walk has cost 1). There is no limit on the number of letters that can fit in a mailbox.

Our agents have received an arbitrary bag of unsorted letters; the agents' sets of letters are disjoint. We would like to enable the agents to negotiate an exchange of letters such that they both lower their final costs.

This domain is inherently "cooperative," meaning that there will always be at least one deal that increases or maintains each agent's "isolated" utility, the utility he could achieve if the other agent did not exist.[1] This paper deals solely with cooperative interactions.

2.2 Domain Definitions

There is a weighted graph $G = G(V, E)$ which is the city map. Each $v \in V$ represents an address, and each $e \in E$ represents a road. There is a special address in V called the "Post Office." The weight function $w: E \to \mathbb{N}$ is the distance of any given road. For each edge $e \in E$, $w(e)$ is the "length" of e, or the "cost" of e. Each agent has a set of letters $L_i: i \in \{A, B\}$ which he has to deliver ($L_A \cap L_B = \emptyset$).

If l is a letter then $\text{Address}(l) \in V$ will be the address of the letter l. If L is a set of letters then $\text{Address}(L)$ will stand for $\{\text{Address}(l): l \in L\} \subseteq V$. $\text{Address}(L_i)$ is the set of all the addresses that agent i has to visit in order to deliver all his letters. If $X \subseteq V$ then $\text{Cost}(X) \in \mathbb{N}$ will be the weight of the minimal weight cycle that starts at the post office, visits all the vertices in X, and ends at the post office. If L is a set of letters then $\text{Cost}(L)$ will be shorthand for $\text{Cost}(\text{Address}(L))$. In order to achieve his goal, agent i will have to walk at least $\text{Cost}(L_i)$ meters.

Theorem 1 *For any two sets of letters L_1, L_2:*

$$\text{Cost}(L_1) + \text{Cost}(L_2) \geq \text{Cost}(L_1 \cup L_2).$$

Proof. Doing the minimal cycle for L_1 and then doing the minimal cycle for L_2 is only one possible way of doing the cycle that delivers $L_1 \cup L_2$, and can not be shorter than the *minimal* cycle that delivers $L_1 \cup L_2$. □

[1] This contrasts with non-cooperative domains, where each agent would do better if it were alone, and will have to lower its utility just to handle interference in the group setting. The aim there is to keep from lowering your utility more than necessary.

2.3 Initial Assumptions

1. **Expected Utility Maximizer:** *Each agent wants to maximize his expected utility.*
2. **Complete Knowledge:** *Each agent knows all relevant information.*
3. **No History:** *There is no consideration given by the agents to the past or future; each negotiation stands alone.*
4. **Commitments are Verifiable:** *If agent i commits to delivering some letter $l \in L_j: i \neq j$ as part of a negotiation agreement, agent j can verify whether i carried out his commitment.*

3 Negotiation with Complete Information

3.1 Definitions

Definition 1 *A* Deal *is a division of $L_A \cup L_B$ to two disjoint subsets, (D_A, D_B) such that $D_A \cup D_B = L_A \cup L_B$, and $D_A \cap D_B = \emptyset$. This deal means that each agent i agrees to deliver all the letters in D_i.*

There may be many possible deals; we want the agents to negotiate so as to agree on a single deal. First we have to decide what constitutes a rational deal, then we have to find a way to make the two agents converge on a single rational deal in a finite negotiation process.

Definition 2 *If (D_A, D_B) is a* Deal, *then*

$$\text{Utility}_i(D_A, D_B) = \text{Cost}(L_i) - \text{Cost}(D_i).$$

In other words, the utility for agent i of a deal is the difference between the cost of achieving his goal alone, and the cost of his part of the Deal.

A deal δ is called *individual rational* if $\forall i \in \{A, B\}$, $\text{Utility}_i(\delta) \geq 0$. Let δ and δ' be two deals. We say that δ *dominates* δ', and write $\delta \succ \delta'$, if and only if $(\text{Utility}_A(\delta), Utility_B(\delta)) \gg (Utility_A(\delta'), Utility_B(\delta'))$.[2] A deal δ is called *pareto optimal* if there does not exist another deal δ' such that $\delta' \succ \delta$ [Roth, 1979; Luce and Raiffa, 1957; Harsanyi, 1977]. The set of all deals that are *individual rational* and *pareto optimal* is called the *negotiation set* (NS) [Harsanyi, 1977]. The Deal $\theta = (L_A, L_B)$ will be called the *conflict* deal. This deal is a conflict because no agent will agree to deliver any letters other than his own.

Theorem 2 *For any G, w, L_A, and L_B,* NS *is not empty.*

Proof. For the proof of this theorem and subsequent theorems, see [Zlotkin, 1988].

[2] It is better for at least one agent and not worse for the other; for vectors α and β, $\alpha \gg \beta$ if and only if $\forall i(\alpha_i \geq \beta_i)$, and $\exists j(\alpha_j > \beta_j)$.

3.2 The Negotiation Protocol

In this section we present a negotiation protocol that ensures convergence to a single deal in NS in a finite negotiation process. The negotiation protocol is iterative: at each step both agents offer (simultaneously) a deal from NS. In each step at least one of the agents has to make a concession, otherwise they reach a conflict.

We will be making use of the function π; if δ is a deal then $\pi(\delta)$ is the product of the two agents' utilities from δ.

The protocol is as follows. In each step $t \geq 0$, both agents simultaneously offer the deals $\delta(A,t)$ and $\delta(B,t)$, such that both are in NS and $\forall i \in \{A,B\}$, $\forall t > 0$, $\text{Utility}_i(\delta(i,t)) \leq \text{Utility}_i(\delta(i,t-1))$. The negotiation can end in one of two ways. We have *conflict* at step t if $\forall i \in \{A,B\}$, $\text{Utility}_i(\delta(i,t)) = \text{Utility}_i(\delta(i,t-1))$, in which case they then agree on the conflict deal θ. We have *agreement* at step t if $\exists j \neq i \in \{A,B\}$ such that $\text{Utility}_j(\delta(i,t)) \geq \text{Utility}_j(\delta(j,t))$. If it is true *only* for $j = A$, then they agree on the deal $\delta(B,t)$. If it is true *only* for $j = B$, then they agree on the deal $\delta(A,t)$. If it is true for *both* $j = A$ and $j = B$, then they will agree on the deal $\delta(k,t)$ such that $\pi(\delta(k,t)) = \max\{\pi(\delta(A,t)), \pi(\delta(B,t))\}$. If it is true for *both* $j = A$ and $j = B$, and $\pi(\delta(A,t)) = \pi(\delta(B,t))$, then they have to flip a coin and choose between the deals $\delta(A,t)$ and $\delta(B,t)$.[3]

Theorem 3 *Using this protocol, the two agents will reach an agreement on a deal after a finite number of steps.*

3.3 Negotiation Strategies

It is clear that the agents using this protocol can run into a conflict. However, if the conflict deal θ is not in NS it would be *irrational* to run into a conflict.

Definition 3 *A* negotiation strategy *is a function from the history of the negotiation to the current message (offer) that is consistent with the negotiation protocol.*

What will be a rational negotiation strategy? If after step t, agent A decides not to make a concession, he takes a risk that agent B will also not make a concession, and they will run into a conflict. Let $\delta_i = \delta(i,t)$ for $i \in \{A,B\}$, and assume that t is not the last step of the negotiation, meaning that $\forall i \neq j \in \{A,B\}$, $\text{Utility}_i(\delta_j) < Utility_i(\delta_i)$. Let p_B^A be the subjective probability that player A associates with the possibility that player B will firmly stick to his own last offer δ_B and will not make further concessions. By Assumption 1, which states that agent A wants to maximize his expected utility, he will stick to his own last offer δ_A only if $(1-p_B^A)\text{Utility}_A(\delta_A) \geq \text{Utility}_A(\delta_B)$, that is, if

$$p_B^A \leq \frac{\text{Utility}_A(\delta_A) - \text{Utility}_A(\delta_B)}{\text{Utility}_A(\delta_A)}.$$

For each step t, and for each agent i, we will define the function $\text{Risk}(i,t)$ to be:

$$\begin{cases} 1 & \text{if } \text{Utility}_i(\delta(i,t)) = 0 \\ \frac{\text{Utility}_i(\delta(i,t)) - \text{Utility}_i(\delta(j,t))}{\text{Utility}_i(\delta(i,t))} & \text{otherwise} \end{cases}$$

[3] Later, we will call this last possibility a "mixed deal."

In other words,

$$\text{Risk}(A,t) = \frac{\text{the utility A loses by accepting B's offer}}{\text{the utility A loses by causing a conflict}}.$$

If t is not the last step in the negotiation protocol, then $\forall i \in \{A,B\}$, $0 \leq \text{Risk}(i,t) \leq 1$; otherwise, $\exists i \in \{A,B\}$ such that $\text{Risk}(i,t) \leq 0$.

$\text{Risk}(A,t)$ is an indication as to how much A is willing to risk a conflict by sticking to his last offer. As $\text{Risk}(A,t)$ grows, agent A has less to lose from a conflict, and he will be more willing to take the risk of reaching one. It is irrational for A to stick to his last offer if $\text{Risk}(A,t) < \text{Risk}(B,t)$.

Definition 4 *Agent A will be said to be using a* rational negotiation strategy *if at any step $t+1$ that A sticks to his last offer,* $\text{Risk}(A,t) > \text{Risk}(B,t)$.

Definition 5 *$SC(A,t)$ is the set of all the* sufficient concessions *of A at step t. That is, if agent A offers in the "next step" a deal from $SC(A,t)$, then if B does not make a concession in that same "next step," B will have to do so in the step after that (assuming that B is using a rational strategy). δ^* will be a* minimal sufficient concession *of A in step t if* $\text{Utility}_B(\delta^*) = \min_{\delta \in SC(A,t)} \text{Utility}_B(\delta)$.

Zeuthen Strategy: Agent A starts the negotiation by offering B the *minimal offer*, meaning that $\text{Utility}_B(\delta(A,1)) = \min_{\delta \in \text{NS}}\{\text{Utility}_B(\delta)\}$. Agent A will make a *minimal sufficient concession* at step $t+1$ if and only if $\text{Risk}(A,t) \leq \text{Risk}(B,t)$.

Theorem 4 *(***Harsanyi***) If both agents are using Zeuthen strategies, they will agree on a deal $\delta^* \in$ NS, such that $\pi(\delta^*) = \max_{\delta \in \text{NS}}\{\pi(\delta)\}$*

3.4 Equilibrium

Definition 6 (Nash) *A negotiation strategy s will be in* equilibrium *if the following condition holds: under the assumption that A uses s, B prefers s to any other strategy.*

This is of particular interest to the designers of automated agents, since to be really useful, agent strategies should be in equilibrium—it can be publicly known that an agent is using a particular strategy, and if that strategy is in equilibrium, no other designer can take advantage of that agent by choosing a different strategy. In essence, the designers are involved in a meta-game in choosing their agents' strategies.

The Zeuthen strategy is *not* in equilibrium, because under the assumption that agent A is using the Zeuthen strategy, agent B can increase his utility by *not* making a concession in step $t+1$, such that $\text{Risk}(i,t) = \text{Risk}(j,t)$ and $t+1$ is going to be the last step. If *either* agent makes the minimal sufficient concession at the last step, they reach an agreement.

Example: Consider the case where the two agents each get a letter l_i, and both letters are addressed to the same address $\text{Address}(l_A) = \text{Address}(l_B)$. There are only two deals in NS: $(\{l_A, l_B\}, \emptyset)$ and $(\emptyset, \{l_A, l_B\})$. If they both use the Zeuthen strategy then in step 1, A offers $(\emptyset, \{l_A, l_B\})$ and B offers $(\{l_A, l_B\}, \emptyset)$. Because

Risk$(A, 1)$ = Risk$(B, 1)$, they both have to concede. In step 2, A offers $(\{l_A, l_B\}, \emptyset)$ and B offers $(\emptyset, \{l_A, l_B\})$. This is the only minimal sufficient concession for both of them. Because $\forall i \neq j \in \{A, B\}$, Utility$_i(\delta(j, 2))$ > Utility$_i(\delta(i, 2))$, and Risk$(A, 2)$ = Risk$(B, 2)$, they have to flip a coin. If B would stick to his first offer in step 2 instead of making a concession, the negotiation would end up with both agents agreeing that A takes both letters with probability 1, instead of choosing which of the two agents will take both letters, each with probability 0.5.

If the next step is going to be the last step of the negotiation, we can view that last step as a game in normal form (see Figure 1).[4] δ_i is the minimal sufficient concession of i, and $\forall i \neq j \in \{a, b\}$, $i^{\downarrow}$ is Utility$_i(\delta_i)$, while $i^{\uparrow}$ is Utility$_j(\delta_i)$.

A \ **B**	*I*	*II*
I	0 0	$b^{\downarrow}$ $b^{\uparrow}$
II	$a^{\uparrow}$ $a^{\downarrow}$	$\frac{b^{\downarrow}+a^{\uparrow}}{2}$ $\frac{b^{\uparrow}+a^{\downarrow}}{2}$

II means concede, *I* means do not concede.

Figure 1: The Last Step as a Game in Normal Form

The only equilibrium point is the mixed strategies, $\left(\frac{a^{\downarrow}-b^{\uparrow}}{a^{\downarrow}+b^{\uparrow}}, \frac{2b^{\uparrow}}{a^{\downarrow}+b^{\uparrow}}\right)$ for A and $\left(\frac{a^{\uparrow}-b^{\downarrow}}{a^{\uparrow}+b^{\downarrow}}, \frac{2b^{\downarrow}}{a^{\uparrow}+b^{\downarrow}}\right)$ for B. The expected payoff (utility) if they both play the equilibrium strategy is $\frac{2a^{\downarrow}b^{\uparrow}}{a^{\downarrow}+b^{\uparrow}}$ for A and $\frac{2a^{\uparrow}b^{\downarrow}}{a^{\uparrow}+b^{\downarrow}}$ for B.

Definition 7 *The* extended Zeuthen strategy *will be the Zeuthen strategy, plus the "last step equilibrium strategy" in last step situations.*

Theorem 5 *The Extended Zeuthen Strategy is in equilibrium.*

This, then, is an equilibrium negotiation strategy that allows our agents to negotiate on the discrete task exchanging domain, and reach a rational agreement in a finite negotiation process.

4 Negotiation on Mixed Deals

In the negotiation protocol that we offered in Section 3.2, the agents could reach a point in their encounter where they have to "flip a coin": they will agree on a pair of symmetric deals $(D_A, D_B), (D_B, D_A)$, only one of which they will actually carry out (according to the flip of a coin). This will be called a *mixed deal* (previous deals without the element of probability will be called "pure deals").

[4]This type of interaction is sometimes called "The Game of Chicken" in the game theory literature [Rapoport and Guyer, 1966].

Definition 8 *If (D_A, D_B) is a deal and $0 \leq p \leq 1, p \in$ IR, then $[(D_A, D_B){:}\, p]$ will be a* mixed deal. *The meaning of such a deal is that the agents will perform (D_A, D_B) with probability p, or (D_B, D_A) with probability $1 - p$.*

In the negotiation protocol of Section 3.2, agents might agree in the end on a mixed deal, but they had to offer only pure deals during every step of the negotiation. What would happen if we allowed the agents to offer mixed deals as well?

Definition 9 *If $[(D_A, D_B){:}\, p]$ is a mixed deal then* $\text{Cost}_i([(D_A, D_B){:}\, p]) = p\,\text{Cost}(D_i) + (1-p)\text{Cost}(D_j)$. *If $[\delta{:}\, p]$ is a mixed deal then* $\text{Utility}_i([\delta{:}\, p]) = \text{Cost}(L_i) - \text{Cost}_i([\delta{:}\, p])$.

The definitions of domination between two mixed deals, individual rational, pareto optimal, and NS are equivalent to the case of pure deals.

Theorem 6 *There exists a number $0 \leq p \leq 1, p \in$ IR, such that the mixed deal $m = [(\{L_A \cup L_B\}, \emptyset){:}\, p] \in$ NS and $\pi(m) = \max_{d \in \text{NS}} \pi(d)$. This deal will be called the "all-or-nothing" deal.*

We see, somewhat surprisingly, that the agents can always agree on a deal in which one of them does all the work with some probability. From an expected utility point of view, this deal is as good as any other deal, but deals that divide the letters between the two agents may take less *time* to execute. If we change the definition of the utility function to include time, then the negotiation set may be changed. Of course, if we assume that an agent can go home after he has done his part of the deal, the negotiation set will not change.[5] If we, however, assume that each agent has to wait at the post office until all letters have been delivered, then the negotiation set will be totally different: even the deal where agent A delivers all the letters may not be good for agent B, who has to wait at the post office. B might prefer some other deal in which he delivers some letters but uses less total time.

Theorem 7 *The agents using the Zeuthen strategy will agree on a mixed deal d such that* Utility$_A(d)$ = Utility$_B(d)$.

Theorem 8 *In the case of mixed deals the Zeuthen strategy is in equilibrium.*

5 Negotiation with Incomplete Information

In Section 2.3 we assumed complete knowledge of all relevant information. Only under this assumption can our agents use the negotiation protocol that we offered in Section 3.2.[6] The subject of this section is what the two agents can do when this information is not available to them.

[5]This is because the time it takes to deliver the letters in his part of the deal is a linear function of the distance he has to walk, and if the utility function is linear then the negotiation set will not change.

[6]The reason for this is because the protocol involves making offers only from NS, which in turn requires the agents to know their opponent's goals (set of letters).

Let us assume that G and w are common knowledge and that each agent i knows L_i, but does not necessarily know $L_j: j \neq i$. What can the agents do in this situation?

A trivial solution would be the mutual exchange of missing information at the beginning of the negotiation. The agents, acting as if their new information were true, then continue negotiating as in Section 3.2. This means that we will add a "-1 phase" to the negotiation in which both agents simultaneously broadcast L_A^* and L_B^*. Can we offer a "good" strategy to play this "-1 phase" game? We would like to convince the agents to tell the truth,[7] and we are even willing to introduce a penalty (against an agent that is proven to have lied) in order to encourage agents to tell the truth.

If the cost of the penalty against a (discovered) lying agent is infinity, then if there is *any* positive probability of being discovered, it will be irrational for an agent to lie. If all lies *might* be discovered, even with a very small probability, then the strategy in the "-1 phase" game of "telling the truth" is in equilibrium.

There are actually two different kinds of possible lies. An agent might broadcast false information about himself, or he might commit himself to doing something as part of an agreement, and then not carry out the commitment. We will assume that the second kind of lie is impossible (Assumption 4), and will thus only concern ourselves with the first kind of lie.

Furthermore, we introduce another assumption:

5. **Discovery During Negotiation:** *False information can be discovered only during the negotiation process, not afterwards.*

Under Assumptions 4 and 5, and if the cost of the penalty for a discovered lie is infinite, is the strategy in a "-1 phase" game of "telling the truth" (i.e., broadcasting the true L_i) in equilibrium? More specifically, assuming that B is going to tell the truth, and that A is completely aware of the encounter's true information (and that B is truthful), can A do better by broadcasting something other than his true L_A? We will consider two typical lies in which A might engage: hiding some letters, and creating some phantom letters.

5.1 Hiding Letters

Because of Assumption 4, and our additional Assumption 5, it turns out that if an agent simply *hides* a letter during a negotiation, it is a "safe" lie—it will never be discovered, and no penalty will be levied against the lying agent. It may in fact help the agent if he lies.

Example: Let the graph be as in Figure 2; the length of each edge is 1. The post office is at node a. Agent B must deliver a letter β_e to node e, while agent A must deliver letters α_b and α_f to nodes b and f. Notice that the graph has a cycle, and that each agent needs to minimally travel a distance equal to the length of the cycle in order to deliver his own letters. Even though agent B, for example, need not actually travel the entire cycle (since he could backtrack after visiting e), he will be indifferent between backtracking and completing the cycle, since they cost the same.

Figure 2: Hiding Letter Example

5.1.1 Pure Deals

If both agents tell the truth (and use some Zeuthen Strategy), they will end the negotiation on pure deals agreeing on the mixed deal $[(\{\alpha_b, \beta_e, \alpha_f\}, \emptyset) : \frac{1}{2}]$ and the expected utility for A and B is 4.

What happens when A "hides" α_b, and tells B that $L_A = \{\alpha_f\}$?

The only deal in NS would be $(\emptyset, \{\beta_e, \alpha_f\})$. This is because it would not be individual rational for A to visit e; B will thus *have* to visit e, and it would not be pareto optimal if he doesn't deliver A's letter to f on the way. B will have to agree to take A's letter to f, and meanwhile A can go to b and deliver his hidden letter. A's expected utility will be 6, instead of 4 (if he were to tell the truth), because, under Assumption 5, there is no possibility that A's lie will be discovered.

5.1.2 Mixed deals

If the agents negotiate on mixed deals, and they both tell the truth, they will reach the same agreement that they would have in the pure deals case. In the situation where A is hiding α_b, they will agree on $[(\{\beta_e, \alpha_f\}, \emptyset) : \frac{3}{8}]$[8] and the expected cost to A will be $\frac{3}{8} \cdot 8 + \frac{5}{8} \cdot 2 = 4\frac{1}{4}$, so the expected utility for A will be $8 - 4\frac{1}{4} = 3\frac{3}{4}$, which is less than 4. Thus, this lie will not help A in this case.

5.2 Phantom letters

Example: Consider the graph on the left of Figure 3 (the length of each edge is written next to it). L_A is $\{\alpha_a, \alpha_c\}$, and L_B is $\{\beta_b, \beta_c\}$.

5.2.1 Pure Deals

If both agents tell the truth (and use some Zeuthen Strategy), they will end the negotiation on pure deals agreeing on the mixed deal $[(\{\alpha_a, \beta_b\}, \{\beta_c, \alpha_c\}) : \frac{1}{2}]$.

What happens when A creates a phantom letter, and tells B that he has another letter (α_d) to deliver to node d (see the graph on the right of Figure 3)? It would not be individual rational for B to visit d; A will thus "have" to visit d, and he could deliver B's letter to c on his way.

[7] For one thing, telling the truth is the cheapest alternative (from a computational point of view) in our model of encounters.

[8] This is the only deal in NS that satisfies Theorem 7.

Figure 3: Phantom Letter Example

B will agree to deliver the letters to a and b, and A will go only until he reaches node c—he has nothing to do at node d. A will get an expected utility of 4, instead of the 3 he would get if he told the truth. This lie is also a "safe" lie, since it is a private action; under our Assumption 4, B cannot verify whether this letter was delivered.

5.2.2 Mixed Deals

The lie of creating a phantom letter is not "safe" if the agents are negotiating on mixed deals. There is always a positive probability that they will agree on a deal where B delivers the phantom letter (because of the "all-or-nothing" deal result in Theorem 6), so there is a positive probability that this lie will be discovered. If B will have to deliver that particular letter, then he will ask A to give him this phantom letter (which does not exist).

As we can see from the case of negotiations on pure deals, there exist some situations where beneficial safe lies can be found. However, we do have the following theorem which states that under some circumstances, there are no beneficial safe lies.

Theorem 9 *Under the assumption that the two agents will always agree on the "all-or-nothing" deal, when negotiating on mixed deals there never exists a beneficial safe lie.*

Our conjecture is that this theorem is true even without the assumption that the agreement will always be the "all-or-nothing-deal," but the proof of the conjecture remains for future work.

Under Assumptions 4 and 5, and an infinite penalty cost for discovered lies, the strategy "tell the truth" for the "-1 phase" game is in equilibrium when the negotiation is on mixed deals, with incomplete information.

6 Conclusion

In order to design agents with sophisticated interaction capabilities, we must first develop sufficiently powerful models for analyzing and modeling these capabilities. In the case of negotiation, game theory offers a starting point for the development of this formal model.

We have here analyzed negotiation among agents in a cooperative, discrete domain. A novel negotiation protocol was introduced that allows agents to agree on rational (pareto optimal, utility maximizing) deals. Moreover, we offered several negotiation strategies that are in equilibrium, meaning that if it were commonly known that the said strategy was being built into automated agents, no one could benefit by choosing a *different* strategy for their own agent.

The case of lying among agents who have incomplete information was also considered. It was shown that on pure deals in discrete domains, there may exist beneficial, safe lies. However, when negotiating on mixed deals (and thus operating in a continuous domain) using our protocol, there will never exist a beneficial, safe lie for any agent, even if one agent has complete knowledge of the other.

References

[Durfee, 1988] Edmund H. Durfee. *Coordination of Distributed Problem Solvers.* Kluwer Academic Publishers, Boston, 1988.

[Harsanyi, 1977] John C. Harsanyi. *Rational Behavior and Bargaining Equilibrium in Games and Social Situations.* Cambridge University Press, Cambridge, 1977.

[Luce and Raiffa, 1957] R. Duncan Luce and Howard Raiffa. *Games and Decisions.* John Wiley & Sons, Inc., New York, 1957.

[Malone *et al.*, 1988] Thomas W. Malone, Richard E. Fikes, and M. T. Howard. Enterprise: A market-like task scheduler for distributed computing environments. In B. A. Huberman, editor, *The Ecology of Computation.* North-Holland Publishing Company, Amsterdam, 1988.

[Rapoport and Guyer, 1966] Anatol Rapoport and M. Guyer. A taxonomy of 2×2 games. *Yearbook of the Society for General Systems Research*, XI:203–214, 1966.

[Rosenschein and Genesereth, 1985] Jeffrey S. Rosenschein and Michael R. Genesereth. Deals among rational agents. In *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, pages 91–99, Los Angeles, California, August 1985.

[Roth, 1979] Alvin E. Roth. *Axiomatic Models of Bargaining.* Springer-Verlag, Berlin, 1979.

[Smith, 1978] Reid G. Smith. *A Framework for Problem Solving in a Distributed Processing Environment.* PhD thesis, Stanford University, 1978.

[Sycara, 1988] Katia P. Sycara. Resolving goal conflicts via negotiation. In *Proceedings of the Seventh National Conference on Artificial Intelligence*, pages 245–250, St. Paul, Minnesota, August 1988.

[Zlotkin, 1988] Gilad Zlotkin. Negotiation and cooperation among autonomous rational agents. Master's thesis, Hebrew University, 1988.